

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

03500.017558



PATENT APPLICATION

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of:)
HIROKI KISHI) : Examiner: Unassigned
Application No.: 10/656,375) : Group Art Unit: 2613
Filed: September 8, 2003) :
For: IMAGE PROCESSING) :
APPARATUS AND METHOD) :
THEREOF) : May 19, 2004

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

SUBMISSION OF PRIORITY DOCUMENT

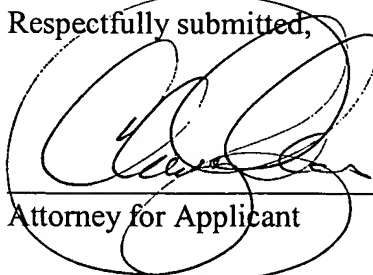
Sir:

In support of Applicant's claim for priority under 35 U.S.C. § 119, enclosed is
a certified copy of the following foreign application:

2002-273254, filed September 19, 2002

Applicant's undersigned attorney may be reached in our Washington, D.C. office by telephone at (202) 530-1010. All correspondence should continue to be directed to our address given below.

Respectfully submitted,

A handwritten signature in dark ink, appearing to be "C. Cella", is written over a horizontal line. The signature is enclosed within a large, loopy circular flourish.

Attorney for Applicant

Registration No. 32,078

FITZPATRICK, CELLA, HARPER & SCINTO
30 Rockefeller Plaza
New York, New York 10112-3800
Facsimile: (212) 218-2200
CPW\gmc

DC_MAIN 166610v1

日本国特許庁

JAPAN PATENT OFFICE

Appln. No. 10/656,375^{US/AS}
Filed - 09-08-03
Group - 2613
Hiroki Kishi

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出願年月日
Date of Application: 2002年 9月19日

出願番号
Application Number: 特願2002-273254
[ST. 10/C]: [JP 2002-273254]

出願人
Applicant(s): キヤノン株式会社

2003年10月 7日

特許庁長官
Commissioner,
Japan Patent Office

今井 康



【書類名】 特許願

【整理番号】 4657004

【提出日】 平成14年 9月19日

【あて先】 特許庁長官 太田 信一郎 殿

【国際特許分類】 H04N 5/00

【発明の名称】 画像処理装置及び画像処理方法

【請求項の数】 16

【発明者】

 【住所又は居所】 東京都大田区下丸子3丁目30番2号キャノン株式会社
 内

 【氏名】 岸 裕樹

【特許出願人】

 【識別番号】 000001007

 【住所又は居所】 東京都大田区下丸子3丁目30番2号

 【氏名又は名称】 キャノン株式会社

 【代表者】 御手洗 富士夫

 【電話番号】 03-3758-2111

【代理人】

 【識別番号】 100090538

 【住所又は居所】 東京都大田区下丸子3丁目30番2号キャノン株式会社
 内

 【弁理士】

 【氏名又は名称】 西山 恵三

 【電話番号】 03-3758-2111

【選任した代理人】

【識別番号】 100096965

【住所又は居所】 東京都大田区下丸子 3 丁目 3 0 番 2 号キャノン株式会
社内

【弁理士】

【氏名又は名称】 内尾 裕一

【電話番号】 03-3758-2111

【手数料の表示】

【予納台帳番号】 011224

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 9908388

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 画像処理装置及び画像処理方法

【特許請求の範囲】

【請求項 1】 動画像を符号化する画像処理装置において、
入力した画像データを符号化する画像符号化手段と、
前記画像データとともに入力した音声データを符号化する音声符号化手段と、
前記画像データで表される画像の一部領域を高画質に符号化するように前記画像符号化手段を設定する画像符号化設定手段と、

前記画像符号化設定手段の設定に従って、前記音声データを高品位に処理するように前記音声符号化手段を設定する音声符号化設定手段とを備えたことを特徴とする画像処理装置。

【請求項 2】 請求項 1 において、前記画像符号化設定手段は、前記画像データからなる動画像の任意の n 枚 (n は 1 以上の整数) のフレーム画に対して選択的に一部領域を高画質にする設定が可能であって、前記音声符号化設定手段は前記画像符号化設定手段によって設定が行われた前記フレーム画の期間に相当する音声データを高品位に処理するよう設定することを特徴とする画像処理装置。

【請求項 3】 請求項 1 において、前記音声符号化手段は前記音声符号化設定手段の設定に応じて、前記音声データを高品位に処理する期間は、高品位に処理されない他の期間より、割り当てる符号量を多くして符号化することを特徴とする画像処理装置。

【請求項 4】 請求項 1 において、前記音声符号化手段は前記音声符号化設定手段の設定に応じて、前記音声データを高品位に処理する期間は、前記音声データの符号化を実行しないことを特徴とする画像処理装置。

【請求項 5】 請求項 1 において、前記音声符号化手段は複数の音声符号化回路を備え、前記音声符号化設定手段の設定に応じて、前記音声データを高品位に処理する期間は、入力された音声データに応じて前記複数の音声符号化回路を適応的に切り替えることを特徴とする画像処理装置。

【請求項 6】 請求項 1 において、前記画像符号化設定手段はユーザーの指示に応じて、ユーザーが指定した領域を高画質に符号化するように設定すること

を特徴とする画像処理装置。

【請求項 7】 請求項 1 において、前記画像符号化設定手段は画像の重要度を指定するユーザーの指示に応じて、前記画像の一部領域を高画質に符号化するように設定することを特徴とする画像処理装置。

【請求項 8】 請求項 6 または 7 において、前記画像符号化設定手段は R O I を設定し、前記画像符号化手段は R O I 符号化を実行することを特徴とする画像処理装置。

【請求項 9】 動画像を入力してその画像データを符号化する画像符号化ステップと、

前記画像データとともに入力した音声データを符号化する音声符号化ステップと、

前記動画像を構成する各画面の一部領域を高画質に符号化するように前記画像符号化ステップを設定する画像符号化設定ステップと、

前記画像符号化設定ステップでの設定に従って、前記音声データを高品位に処理するように前記音声符号化ステップを設定する音声符号化設定ステップとを有することを特徴とする画像処理方法。

【請求項 10】 請求項 9 において、前記画像符号化設定ステップは、前記動画像の任意の n 枚 (n は 1 以上の整数) のフレーム画に対して選択的に一部領域を高画質にする設定が可能であって、前記音声符号化設定ステップは前記画像符号化設定ステップによって設定が行われた前記フレーム画の期間に相当する音声データを高品位にするよう設定することを特徴とする画像処理方法。

【請求項 11】 請求項 9 において、前記音声符号化ステップは前記音声符号化設定ステップの設定に応じて、前記音声データを高品位に処理する期間は、高品位に処理されない他の期間より割り当てる符号量を多くして符号化することを特徴とする画像処理方法。

【請求項 12】 請求項 9 において、前記音声符号化ステップは前記音声符号化設定ステップの設定に応じて、前記音声データを高品位に処理する期間は、前記音声データの符号化を実行しないことを特徴とする画像処理方法。

【請求項 13】 請求項 9 において、前記音声符号化ステップは複数種の音

声符号化工程を有し、前記音声符号化設定ステップの設定に応じて、前記音声データを高品位に処理する期間は、入力された音声データに応じて前記複数種の音声符号化工程を適応的に切り替えることを特徴とする画像処理方法。

【請求項 14】 請求項 9 において、前記画像符号化設定ステップはユーザーの指示に応じて、ユーザーが指定した領域を高画質に符号化するように設定することを特徴とする画像処理方法。

【請求項 15】 請求項 9 において、前記画像符号化設定ステップは画像の重要度を指定するユーザーの指示に応じて、前記画像の一部領域を高画質に符号化するように設定することを特徴とする画像処理方法。

【請求項 16】 請求項 14 または 15 において、前記画像符号化設定ステップは R O I を設定し、前記画像符号化ステップは R O I 符号化を実行することを特徴とする画像処理方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、画像及び音声を符号化する画像処理装置及びその方法に関するものである。

【0002】

【従来の技術】

従来、画像を記録再生できる記録再生装置として、民生用のデジタルビデオカメラやデジタルスチルカメラがある。

【0003】

これらの装置においては静止画や動画といった画像を記録や伝送する際に、画像データの圧縮処理を行っている。

【0004】

画像の圧縮技術（規格）としては、静止画は J P E G、動画は M o t i o n - J P E G、M P E G 1、M P E G 2 や M P E G 4 といったものが一般的であり、最近では新たに J P E G 2 0 0 0 といった規格も検討されている。

【0005】

また、一部の画像圧縮技術の中には、画像の特定領域を他の領域に比べて高画質になる様に符号化できるものがある。例えば、ROI (Region Of Interest) 符号化法を用いれば、復号後に、ある画面における興味ある領域 (ROI) を高画質に表示、もしくは伝送の初期段階で優先的に表示することが可能になる。

【0006】

このようなROI符号化を用いたカメラとして、撮像画像中の特定領域をROIに設定して圧縮し、記録再生できるデジタルスチルカメラが考えられている (例えば、特許文献1参照。)。カメラでROIとして設定する場合は、その設定される範囲或いは被写体が画面の中で重要な、注目すべき対象物を含んでいる場合が多い。

【0007】

【特許文献1】

特開2001-230947号公報

【0008】

【発明が解決しようとする課題】

ところで、動画においてROI符号化を適用した場合には、動画を構成する各フレーム単位でのROI設定が可能であるが、このときROIが含まれているフレーム (ROIフレーム) は注目すべき対象物を含んでいるので、ROIフレームはROIが含まれていないフレーム (非ROIフレーム) より、撮像者にとって重要となる場合が考えられる。

【0009】

しかしながら、ROIフレームを含んで符号化されている動画データを再生する際には、符号化された動画データを復号して表示するのみであって、フレーム間での重要度に差が生じている場合であっても、そのフレームの重要度に合わせて、動画と共に再生される音声 (オーディオ) の音質までを変動させることはできなかった。

【0010】

本発明は上記の如き問題点を解決して、画像の重要度に応じて音声も音質を変

更できるよう設定する画像処理装置及びその方法を提供することを目的とする。

【0011】

【課題を解決するための手段】

斯かる目的を達成する為の手段として、本発明は以下の構成からなる手段を有する。

【0012】

本発明の画像処理装置は、入力した画像データを符号化する画像符号化手段と、前記画像データとともに入力した音声データを符号化する音声符号化手段と、前記画像データで表される画像の一部領域を高画質に符号化するように前記画像符号化手段を設定する画像符号化設定手段と、前記画像符号化設定手段の設定に従って、前記音声データを高品位に処理するように前記音声符号化手段を設定する音声符号化設定手段とを備えたことを特徴とする。

【0013】

また、本発明の画像処理方法は、動画像を入力してその画像データを符号化する画像符号化ステップと、前記画像データとともに入力した音声データを符号化する音声符号化ステップと、前記動画像を構成する各画面の一部領域を高画質に符号化するように前記画像符号化ステップを設定する画像符号化設定ステップと、前記画像符号化設定ステップでの設定に従って、前記音声データを高品位に処理するように前記音声符号化ステップを設定する音声符号化設定ステップとを有することを特徴とする。

【0014】

【発明の実施の形態】

以下、図面を参照しながら本発明の好適な実施の形態を説明する。

【0015】

(第1の実施形態)

本発明で扱う動画像は、図1に示したように、所定の周期（一例としては1/30秒）ごとの画面（フレーム）が複数集まって構成される画像データと、それに対応する音声データから構成されている。

【0016】

本実施の形態においてはこのような動画像のデータ（動画データ）に対して、図3のように、任意のフレームにおける任意の領域に対して、関心領域（ROI）を設定するものとする。さらに、本実施の形態における画像処理装置は、画像のROIが設定されている間（ROI時間）に流れるオーディオ（音声）の音質が、画像のROIが設定されない間（非ROI時間）に流れる音声の音質より高くなるように、ビットレート（割り当てビット量）を変えて音声データを符号化する構成となっている。以下にその構成の詳細を記す。

【0017】

本実施の形態における画像処理装置のブロック図を図2に示し、当該画像処理装置の動作を説明するフローチャートを図4に示す。

【0018】

図2において、画像処理装置200の構成は、画像データ取得部201、フレームデータ符号化部202、音声データ取得部203、音声データ符号化部204、データ統合部205、動画像符号化データ出力部206、ROI設定指示部207、ROI設定部208から成る。

【0019】

図4のフローチャートに従いながら説明する。

【0020】

まず、動画データと音声データが、画像データ取得部201と音声データ取得部203からそれぞれ入力される（ステップS401）。この画像データ取得部201と音声データ取得部203は、例えばデジタルビデオカメラやデジタルスチルカメラ等の撮像装置、或いはCCDなどの撮像デバイスやネットワーク回線のインターフェース及びマイクや外部音声の入力装置等が用いられる。また、画像データ取得部201と音声データ取得部203はRAM、ROM、ハードディスク、CD-ROM等の記録媒体とその再生装置であっても良い。

【0021】

画像データ取得部201から取得された動画データは、フレーム単位のデータ（フレームデータ）毎にフレームデータ符号化部202に入力され、音声データ取得部203から取得された音声データは、音声データ符号化部204に入力さ

れる。

【 0 0 2 2 】

この動画データのフレームデータ符号化部 2 0 2 への入力に際して、ユーザーが R O I 設定指示部 2 0 7 から R O I の設定を指示していない場合（ステップ S 4 0 2 で N o）、フレームデータはフレームデータ符号化部 2 0 2 でフレーム単位毎に通常の符号化が行われ、音声データは音声データ符号化部 2 0 4 で通常どおり低ビットレートでの符号化が行われる（ステップ S 4 0 2）。生成されたフレーム符号化データと音声符号化データは、データ統合部 2 0 2 に出力される。

【 0 0 2 3 】

一方、動画データのフレームデータ符号化部 2 0 2 への入力に際して、ユーザーが R O I 設定指示部 2 0 7 から R O I の設定を指示している場合（ステップ S 4 0 2 で Y e s）、R O I 設定部 2 0 8 がフレーム中のどの領域が R O I であるかを示す情報をフレームデータ符号化部 2 0 2 に送り、R O I の設定が行われる旨の情報を音声データ符号化部 2 0 4 へ送るといった R O I の設定を行う（ステップ S 4 0 4）。その上で、フレームデータ符号化部 2 0 2 は入力したフレームデータを、R O I に指定された領域を高画質に、その他の領域を低画質にするよう符号化（R O I 符号化）して、また音声データ符号化部 2 0 4 は入力された音声データを高音質となるよう高ビットレートで符号化する（ステップ S 4 0 5）。生成されたフレーム符号化データと音声符号化データは、データ統合部 2 0 2 に出力される。

【 0 0 2 4 】

データ統合部 2 0 5 に音声符号化データとフレーム符号化データが入力されると、図 5 に示されているように、フレーム符号化データと音声符号化データが所定の規則に基づき並べられデータの統合がなされる。さらに、図 6 に示されているように、ヘッダ等の復号に必要となるデータが所定の個所に挿入されて、動画像符号化データが生成される（ステップ S 4 0 6）。

【 0 0 2 5 】

その後、生成された動画像符号化データは、動画像符号化データ出力部 2 0 6 から外部へ出力される（ステップ S 4 0 7）。この動画像符号化データ出力部 2

06には、有線、公衆回線、無線回線、LAN等のインターフェースを用いることができ、出力された動画像符号化データは記録装置や表示装置等に伝送される。

【0026】

次にROI設定部208の動作を説明する。図7にROI設定部208の詳細ブロック図を示すと共に、ROI設定に関する処理のフローチャートを図8に示す。

【0027】

ROI設定指示部207がユーザーからROIの設定の指示を受けると、その旨をROI設定部208に伝達する。ROI設定部208がこの伝達を受けると、前もってユーザーが設定しておいたROIオブジェクトがいずれであるかを手する（ステップS801）。

【0028】

このROIオブジェクトとは、画像処理装置200が不図示のメモリ上に格納しているオブジェクトのことであり、ユーザーが設定するROIにはROIオブジェクトが含まれる必要がある。なお、ユーザーがROIオブジェクトを選択する方法の一形態として、図9に示されているように、画像処理装置200を搭載したデジタルビデオカメラや、デジタルスチルカメラに具備されている液晶ディスプレイ上で行う方法が挙げられる。

【0029】

ROIオブジェクト情報の入手が終ると、フレームデータ入力部701はフレームデータ符号化部202から符号化前のフレームデータのコピーを得る（ステップS802）。次に、ROIオブジェクト認識部702は、ユーザーが選択したROIオブジェクトに相当する、当該フレームデータ上のオブジェクトを検索する（ステップS803）。ROI認識部702はこの検索によりオブジェクトを検出すると（ステップS804）、図10に示されているように、このオブジェクトを包含するような領域をROIと設定する（ステップS805）。その後、ROIマスク生成部703は、図11に示されているように、フレームにおけるROIの位置を示すROIマスクを生成し（ステップS806）、このROI

マスクをROIマスク出力部704に出力する。

【0030】

ROIマスク出力部704は、入力されたROIマスク情報をフレームデータ符号化部202に出力する（ステップS807）。

【0031】

その後、続いて処理するフレームがROIフレームであるかどうか判断され（ステップS808）、当該フレームがROIフレームであれば処理がステップS802に戻される。

【0032】

次に、フレームデータ符号化部202の詳細ブロック図を図12に示し、当該フレームデータ符号化部202の処理動作を説明する為のフローチャートを図13に示し、これらの図を用いてフレームデータの符号化の処理を詳細に説明する。

【0033】

本実施の形態におけるフレーム中の符号化対象となるフレームデータは、8ビットのモノクロフレームデータとする。しかしながら、各画素4ビット、10ビット、12ビットといった具合に8ビット以外のビット数で表すモノクロ画像、或いは各画素における各色成分（RGB/Lab/YCrCb）を8ビットで表現するカラーの多値フレームデータである場合に適用することも可能である。また、画像を構成する各画素の状態等を表す多値情報である場合、例えば各画素の色を表す多値のインデックス値である場合にも適用できる。これらに応用する場合には、各種類の多値情報を後述するモノクロフレームデータとすればよい。

【0034】

まず、画像データ取得部201から、フレームデータ/ROIマスク入力部1201へ、符号化対象となるフレームデータを構成する画素データがラスタスキャン順に入力され、離散ウェーブレット変換部1202に出力される（ステップS1301）。

【0035】

また、ROI設定部208が、当該フレームに対応するROIマスクを生成し

ている場合、当処理部からフレームデータ／ROIマスク入力部1201へ、ROIマスクが入力され、バッファ1209へ出力される（ステップS1301）。このフレームデータ／ROIマスク入力部1201は、例えばスキャナ、デジタルカメラ等の撮像装置、或いはCCDなどの撮像デバイス、或いはネットワーク回線のインターフェース等が接続される。また、フレームデータ入力部1201はRAM、ROM、ハードディスク、CD-ROM等の記録媒体から読み出した情報を入力しても良い。

【0036】

離散ウェーブレット変換部1202は、フレームデータ／ROIマスク入力部1201から入力される1つの静止画像中の、1つのタイルデータ $x(n)$ における複数の画素（参照画素）のデータ（参照画素データ）を用いて離散ウェーブレット変換を行う（ステップS1302）。

【0037】

以下に、離散ウェーブレット変換後のフレームデータ（離散ウェーブレット変換係数）を示す。

$$Y(2n) = X(2n) + \text{floor} \{ (Y(2n-1) + Y(2n+1) + 2) / 4 \}$$

$$Y(2n+1) = X(2n+1) - \text{floor} \{ (X(2n) + X(2n+2)) / 2 \}$$

$Y(2n)$, $Y(2n+1)$ は離散ウェーブレット変換係数列であり、 $Y(2n)$ は低周波サブバンド、 $Y(2n+1)$ は高周波サブバンドである。また、上式において $\text{floor} |X|$ は X を超えない最大の整数値を表す。この離散ウェーブレット変換を模式的に表わしたのが図14である。

【0038】

本変換式は一次元のデータに対するものであるが、この変換を水平方向、垂直方向の順に適用して二次元の変換を行うことにより、図15(a)の様なLL, HL, LH, HHの4つのサブバンドに分割することができる。ここで、Lは低周波サブバンド、Hは高周波サブバンドを示している。次にLLサブバンドを、同じ様に4つのサブバンドに分け（図15(b)）、その中のLLサブバンドをまた4サブバンドに分ける（図15(c)）。合計10サブバンドを作る。10個のサブバンドそれぞれに対して、図15(c)の様にHH1, HL1, …と呼

ぶことにする。ここで、各サブバンドの名称における数字を、それぞれのサブバンドのレベルとする。つまり、レベル1のサブバンドは、HL1, HH1, LH1、レベル2のサブバンドは、HL2, HH2, LH2である。なおLLサブバンドは、レベル0のサブバンドとする。LLサブバンドはひとつしかないので添字を付けない。またレベル0からレベルnまでのサブバンドを復号することで得られる復号画像を、レベルnの復号画像と呼ぶ。復号画像は、そのレベルが高い程解像度は高い。

【0039】

10個のサブバンドの変換係数は、いったんバッファ1203に格納されLL, HL1, LH1, HH1, HL2, LH2, HH2, HL3, LH3, HH3の順に、つまり、レベルが低いサブバンドからレベルが高いサブバンドの順に、係数量子化部1204へ出力される。

【0040】

係数量子化部1204では、バッファ1203から出力される各サブバンドの変換係数を各周波数成分毎に定めた量子化ステップで量子化し、量子化後の値（係数量子化値）を生成する（ステップS1303）。係数値をX、この係数の属する周波数成分に対する量子化ステップの値をqとすると、量子化後の係数値Q(X)は次式によって求めるものとする。

$$Q(X) = \text{floor} \{ (X/q) + 0.5 \}$$

【0041】

本実施の形態における各周波数成分と量子化ステップとの対応を図16に示す。同図に示す様に、よりレベルが高いサブバンドの方に、大きい量子化ステップを与えている。なお、各サブバンドの量子化ステップは予め不図示のRAMやROMなどのメモリに格納されているものとする。

【0042】

そして、係数量子化部1204は、現在処理中のフレームデータに付随するROIマスクがバッファ1209に存在する場合（ステップS1304でYes）、これら係数量子化値をシフトアップ部1205へ出力する。また、当該フレームデータに付随するROIマスクがバッファ1209に存在しない場合（ステッ

プS1304でNo)、これら係数量子化値をエントロピー符号化部1206へ出力する。

【0043】

シフトアップ部1205は、バッファ1209に格納されているROIマスクを基に、ROI内の画像の画質が非ROI内の画像の画質より高くなるように、次式に基づき係数量子化値を変更(シフトアップ)する(ステップS1305)。

$$Q' = Q * 2^B; \quad (Q: ROI内に存在する係数量子化値の絶対値)$$

$$Q' = Q; \quad (Q: ROI外に存在する係数量子化値の絶対値)$$

ここで、Bはサブバンド毎に与えられるものであり、注目サブバンドにおいて、各Q'はいかなるQ'よりも大きくなるように設定される。

【0044】

以上の処理により、シフトアップ部1205において指定された空間領域に属する係数量子化値のみがBビット上方にシフトアップされる。

【0045】

図17(a)は各サブバンドにおけるROIと非ROIを示していて、(b)および(c)はシフトアップによる係数量子化値の変化を示したものである。(b)において、3つのサブバンドに各々3個の係数量子化値が存在しており、網がけされた係数量子化値がROIを構成している係数量子化値であり、それらはシフトアップ後(c)のようになる。

【0046】

以上の処理が行われた係数量子化値は、エントロピー符号化部1206へ出力される。

【0047】

エントロピー符号化部1206では、入力された係数量子化値がエントロピー符号化される(ステップS1306)。ここでは、まず、図18に示されているように、入力された係数量子化値の集まりである各サブバンドが矩形(コードブロックと呼ぶ)に分割される。なお、このコードブロックの大きさには、 $2m \times 2n$ (m、nは2以上の整数)等が設定される。

【0048】

さらにこのコードブロックは、図19に示されているように、ビットプレーンに分割される。その上で、図20に示されているように、あるビットプレーンにおける各ビットは、ある分類規則に基づいて3種類に分けられて、同じ種類のビットを集めたコーディングパスが3種類生成される。入力された係数量子化値は、ここで得られたコーディングパスを単位として、エントロピー符号化である二値算術符号化が行われ、エントロピー符号化値が生成される。

【0049】

なお、ここでエントロピー符号化の具体的な処理順序は、1つのコードブロックに注目すると上位ビットプレーンから下位ビットプレーンの順に符号化され、その1コードブロックのあるビットプレーンに注目すると、図20にある3種類のパスを上から順に符号化する様になっている。

【0050】

エントロピー符号化されたコーディングパスは、フレーム符号化データ生成部1207に出力される。

【0051】

フレーム符号化データ生成部1207では、入力された複数のコーディングパスから、単一もしくは複数のレイヤーが構成され、それらレイヤーをデータの単位としてフレーム符号化データが生成される（ステップS1307）。以下にレイヤーの構成に関する説明を行う。

【0052】

当該処理部は、図21に示されているように、複数のサブバンドにおける複数のコードブロックから、エントロピー符号化されたコーディングパスを集めた上で、レイヤーを構成する。なお、図22に示されているように、あるコードブロックからコーディングパスを取得する際、常に該コードブロックにおいて最上位に存在するコーディングパスが選ばれる。

【0053】

その後、フレーム符号化データ生成部1207は、図23に示されているように、生成したレイヤーを、上位に位置するレイヤーから順に並べた上で、その先

頭にヘッダを付加してフレーム符号化データを生成する。このヘッダには、フレームを識別する情報や、当該フレーム符号化データの符号長や入力フレームの縦横のサイズ、圧縮に使用した様々なパラメータ等が格納される。

【0 0 5 4】

このように生成されたフレーム符号化データは、フレーム符号化データ出力部 1 2 0 8 に出力される（ステップ S 1 3 0 8）。

【0 0 5 5】

以上が、フレームデータ符号化部 2 0 2 の詳細な説明である。

【0 0 5 6】

次に、音声データ符号化部 2 0 4 の詳細ブロック図を図 2 4 に示し、当該音声データ符号化部 2 0 4 の処理動作を説明する為のフローチャートを図 2 5 に示し、これらの図を用いて音声データの符号化の処理を詳細に説明する。

【0 0 5 7】

まず、1 6 ビット直線量子化された入力音声データが、音声データ入力部 2 4 0 1 に入力され（ステップ S 2 5 0 1）、サブバンド分割部 2 4 0 2 と高速フーリエ変換部（FFT部） 2 4 0 4 に出力される。

【0 0 5 8】

サブバンド分割部 2 4 0 2 では、入力された音声データは 5 1 2 タップ PFB（ポリフェーズフィルタバンク）により、3 2 個のサブバンドに分割される（ステップ S 2 5 0 2）。以下に、このサブバンド分割の手順を示す。

【0 0 5 9】

まず、5 1 2 個の入力信号サンプル X_0, \dots, X_{511} に対して、

$$Z_i = C_i \times X_i$$

を計算する。ここで、 C_i の一形態として、ISO/IEC 11172-3 内の表で定義されたものが考えられる。

【0 0 6 0】

次に、

【外 1】

$$Y_i = \sum_{j=0}^7 Z_{64j+i}$$

【0 0 6 1】

を算出し、

【0 0 6 2】

【外 2】

$$S_i = \sum_{k=0}^{63} Y_k * \cos \frac{(2i+i)(k-16)\pi}{64}$$

【0 0 6 3】

に従ってサブバンド出力 S_i を導出する。なお本実施の形態において、生成されるサブバンドの数は N 個とする。

【0 0 6 4】

次に、音声データ符号化部 2 0 4 は、不図示のカウンタ i を $i = 0$ に設定し（ステップ S 2 5 0 3）、スケールファクタ抽出部 2 4 0 3 に各サブバンドのスケールファクタの算出を実行させる（ステップ S 2 5 0 4）。このスケールファクタの抽出は、以下のとおり行われる。

【0 0 6 5】

まず、各サブバンドごとに 1 2 サンプルを 1 ブロックとして、絶対値が最大となるサンプルを探索する。さらに、所定のスケールファクタの表から、当該最大となるサンプルの絶対値より大きい最小の値を選び、その値をスケールファクタとする。

【0 0 6 6】

ここで得られたスケールファクタは線形量子化部 2 4 0 7、サイド情報符号化部 2 4 0 8 および心理聴覚モデル部 2 4 0 5 に出力される。また、必要となるスケールファクタが抽出されたサブバンドは、線形量子化部 2 4 0 7 に出力される（ステップ S 2 5 0 5）。

【0 0 6 7】

ここで不図示のカウンタ i が $i = N - 1$ でなければ、 i を 1 増加させた上で処

理をステップ S 2 5 0 3 に戻す。 $i = N - 1$ であれば、処理をステップ S 2 5 0 7 に進める（ステップ S 2 5 0 6）。

【0068】

F F T 部 2 4 0 4 に入力された音声信号は、512 のサンプル数を 1 ブロックとして、M 個のブロックに分割される。その後、各ブロックに対して F F T が行われる（ステップ S 2 5 0 7）。生成された各ブロックの係数値は、心理聴覚モデル部 2 4 0 5 に出力される。

【0069】

心理聴覚分析モデル部 2 4 0 5 では、入力された係数値とスケールファクタを基に、信号対マスク比（SMR）を算出する（ステップ S 2 5 0 8）。この算出された SMR は、動的ビット割り当て部 2 4 0 6 に出力される。なお、この SMR は各サブバンドの音圧と最小マスキングレベルの比で与えられるものである。この SMR の導出は公知であるので、ここではその説明を割愛する。

【0070】

動的ビット割り当て部 2 4 0 6 は、入力された SMR を用いて、各サブバンドに割り当てるビット量を算出して決定する（ステップ S 2 5 0 9）。この算出方法について、図 2 6 のフローチャートを使用し、以下に説明する。

【0071】

まず、各サブバンドのサンプルを入力する（ステップ S 2 6 0 1）。次に、利用可能である総ビット量のうち、ヘッダや利用者が自由に使えるデータ、誤り訂正に必要となるデータ等に必要となるビット量を減算して得られる、割り当て可能ビット量を算出する（ステップ S 2 6 0 2）。

【0072】

次に、信号対雑音比（SMR）や MNR（マスク対雑音比）を使用し、割り当て可能ビット量を基に、各サブバンドに対して量子化ステップを設定し、各サブバンドに割り当てるビット量を算出する（S 2 6 0 3）。この割り当て可能ビット量を導き出す手法は、様々な文献で紹介されているので、この手法に関する説明は割愛される。

【0073】

その後、割り当てられた各サブバンドに対する量子化ステップは、線形量子化部 2 4 0 7 とサイド情報符号化部 2 4 0 8 に出力される（ステップ S 2 6 0 4）。

【0 0 7 4】

なお本実施の形態においては、動的ビット割り当て部 2 4 0 6 は、R O I 時間に流れる音声符号化データの符号量が、非 R O I 時間に流れる音声符号化データの符号量より多くなるように、利用可能である総ビットの割り当てが行われる。これにより、R O I 時間と非 R O I 時間に流れる音声の音質の差を出すことが可能となる。

【0 0 7 5】

動的ビット割り当て決定部 2 4 0 6 により各サブバンドに割り当てられるビット量と量子化ステップが決定されると、線形量子化部 2 4 0 7 は当該量子化ステップを基に量子化を行う（ステップ S 2 5 1 0）。ここで得られた量子化されたサブバンドサンプルは、音声符号化データ形成部 2 4 0 9 に出力される。また、サイド情報符号化部 2 4 0 8 は、入力されたスケールファクタや量子化ステップ情報を符号化し（ステップ S 2 5 1 1）、音声符号化データ形成部 2 4 0 9 に出力する。

【0 0 7 6】

音声符号化データ形成部 2 4 0 9 は、図 2 7 のようにヘッダとサブバンドサンプルとアンシラリデータ（利用者が任意に利用できるデータ）から音声符号化データを形成し（ステップ S 2 5 1 2）、音声符号化データ出力部 2 4 1 0 に出力される。なおヘッダには、当該音声符号化データを復号するのに必要となる、符号化されたサイド情報やサンプリング周波数等の情報が格納される。

【0 0 7 7】

音声符号化データ出力部 2 4 1 0 は、データ統合部 2 0 5 に音声符号化データを出力する（ステップ S 2 5 1 3）。

【0 0 7 8】

以上、本実施の形態では、撮影者にとって重要な時間である R O I 時間において、ビットレート（割り当てビット量）を変更させることで、音声を高音質に符

号化することが可能になる。

【0079】

(第2の実施の形態)

次に実施の形態2について説明する。

【0080】

第1の実施の形態における画像処理装置200では、ROI時間と非ROI時間に流れる音声データの復号音質の差をビットレートの変更により実現した。本実施の形態においては、復号音質の差を、圧縮を行うか否かにより実現する。

【0081】

図31に本実施の形態における画像処理装置3100のブロック図を示す。これは、第1の実施の形態における画像処理装置200に対して、音声データ取得部203の後段にスイッチ3101と、スイッチ3101からデータ統合部205へのパスを挿入したものである。

【0082】

このスイッチ3101は通常、音声データ符号化部204につながっているが、撮像者によりROI設定指示部207からROIの設定の指示が入力されると、スイッチ3301は音声データ符号化部204をバイパスする経路につながる。これにより、音声データの符号化の有無を実現できる。

【0083】

以上、本実施の形態では、非ROI時間に流れる音声データに対して圧縮を行い、ROI時間に流れる音声データに対して圧縮を行わないことにより、2つの時間に流れる復号音声の音質に差をつける方法が示された。

【0084】

(第3の実施の形態)

次に実施の形態3について説明する。

【0085】

実施の形態3における画像処理装置2800の構成を図28に示す。図28の画像処理装置2800において、図2の画像処理装置200の構成と同符号のブロックは図2と同様なので、ここでの説明は書略する。図28に示した本実施の

形態の画像処理装置 2 8 0 0 は、実施の形態 1 における画像処理装置 2 0 0 の音声データ符号化部 2 0 4 を音声データ符号化部 2 8 0 1 に置き換えたものである。

【0 0 8 6】

例えば、画像処理装置 2 8 0 0 を図 9 に記したようなデジタルスチルカメラやデジタルビデオカメラに搭載した場合、撮像中において、時間によって、入力される音声がスピーチのみであるときと、そうでないときがあると考えられる。そこで、画像処理装置 2 8 0 0 は音声データの符号化器として、スピーチデータ用の符号化器と一般音声データ用の符号化器を具備する音声データ符号化部 2 8 0 1 を備えた。これによって、2 つの符号化器が出力する音声符号化データが、復号音声の音質に関して比較された上で、最適な音声符号化データが格納されるようにする。

【0 0 8 7】

しかしこの処理は、高品位に音声データを符号化することが可能であるが、消費電力の増加を招く。そこで本実施の形態においては、特に R O I 時間においてのみ、上述処理を行う構成とする。

【0 0 8 8】

なお、本実施の形態において具備される 2 つの音声データ符号化器は、ビットレートが 9 6 k b p s 以上であると、様々な音声に対して高効率に圧縮する A A C (Advanced Audio Coding) と、6 4 k b p s 以下のビットレートにおけるスピーチデータの圧縮では、A A C に符号化効率で勝る C E L P (Code Excited Linear Prediction) である。

【0 0 8 9】

本実施の形態におけるフレームデータの符号化方法は、実施の形態 1 におけるフレームデータの符号化方法と同様であるので、ここでは、フレームデータの符号化方法に関する説明を割愛する。

【0 0 9 0】

本実施の形態における音声データ符号化部 2 8 0 1 のブロック図を図 2 9 に示す。本実施の形態における音声データの符号化は、3 2 - 6 4 k b p s 程度のビ

ットレートで行われるものとする。

【0091】

まず、フレームデータがROIを設定されずに符号化されている場合、入力された音声データはAAC符号化部2902に流れ、音声データはAAC符号化部により32kbp/s程度のビットレートで符号化される。なお、このAACによる符号化方法は公知であるので、ここではその説明を割愛する。

【0092】

さらに、フレームデータが、ROIを設定された上で符号化されている場合、音声データはAAC符号化部2902とCELP符号化部2903に入力される。当該2つの符号化部は、それぞれ音声データを符号化し、音声符号化データを生成し、最適音声符号化データ決定部2904に出力する。

【0093】

続いて、最適音声符号化データ決定部2904の動作を説明するにあたり、図30に示されている最適音声符号化データ決定部2904の詳細ブロック図を用いる。

【0094】

まず、所定の時間間隔における、AAC符号化部2902で生成された音声符号化データ（AAC音声符号化データ）とCELP符号化部2903で生成された音声符号化データ（CELP音声符号化データ）は、それぞれAAC音声符号化データ入力部3001，CELP音声符号化データ入力部3002から入力され、いったんバッファ3003に蓄積される。

【0095】

AAC復号部3004とCELP復号部3005は、バッファ3003に蓄積されているAAC音声符号化データとCELP音声符号化データを、内部に保持しているメモリにコピーする。その上で、当該2つの復号部は音声符号化データを復号し、復号データを判定部3006に出力する。なお、AACの復号方法とCELPの復号方法は公知であるので、ここに詳細を記さない。

【0096】

判定部3006は、所定の方法に基づき、2つの復号データから得られる復号

音声のどちらが高音質であるかを判定し、低音質な音声符号化データ的一方をバッファ3003上から除去し、高音質な音声符号化データをバッファ3003から音声符号化データ出力部2905に出力する。

【0097】

選択された音声符号化データは、音声符号化データ出力部2905からデータ統合部205に出力される。

【0098】

以上、本実施の形態では、ROI時間にのみ、スピーチ用の音声符号化器と非スピーチ用の音声符号化器の使用を判断しながら使う方法を示した。こうすることで、ROI時間における復号音声の音質を、非ROI時間における復号音声の音質より高めることが可能となる。

【0099】

(その他の実施形態)

全ての実施形態において、フレームデータの系列変換に離散ウェーブレット変換を使用したか、本発明はこれに限定されるわけではない。系列変換に離散コサイン変換等を使用しても構わない。

【0100】

なお、本発明は複数の機器（例えばホストコンピュータ、インターフェース機器、リーダ、プリンタ等）から構成されるシステムの一部として適用しても、1つの機器（例えば複写機、デジタルカメラ、デジタルビデオカメラ等）からなる装置の一部に搭載して適用しても良い。

【0101】

また、本発明は上記実施の形態を実現するための装置及び方法のみに限定されるものではなく、上記システム又は装置内のコンピュータ（CPU或いはMPU）に、上記実施の形態を実現するためのソフトウェアのプログラムコードを供給し、このプログラムコードに従って上記システム或いは装置のコンピュータが上記各種デバイスを動作させることにより上記実施の形態を実現する場合も本発明の範疇に含まれる。

【0102】

またこの場合、前記ソフトウェアに関するプログラムコード自体が上記実施の形態の機能を実現することになり、そのプログラムコード自体、及びそのプログラムコードをコンピュータに供給するための手段、具体的には上記プログラムコードを格納した記憶媒体は本発明の範疇に含まれる。

【0103】

この様なプログラムコードを格納する記憶媒体としては、例えばフロッピー（R）ディスク、ハードディスク、光ディスク、光磁気ディスク、CD-ROM、磁気テープ、不揮発性のメモ리카ード、ROM等を用いることができる。

【0104】

また、上記コンピュータが、供給されたプログラムコードのみに従って各種デバイスを制御することにより、上記実施の形態の機能が実現される場合だけではなく、上記プログラムコードがコンピュータ上で稼動しているOS（オペレーティングシステム）、或いは他のアプリケーションソフト等と共同して上記実施の形態が実現される場合にもかかるプログラムコードは本発明の範疇に含まれる。

【0105】

更に、この供給されたプログラムコードが、コンピュータの機能拡張ボードやコンピュータに接続された機能拡張ユニットに備わるメモリに格納された後、そのプログラムコードの指示に基づいてその機能拡張ボードや機能拡張ユニットに備わるCPU等が実際の処理の一部または全部を行い、その処理によって上記実施の形態が実現される場合も本発明の範疇に含まれる。

【0106】

【発明の効果】

以上説明したように本発明によれば、ROI等の設定によって動画のフレーム間で重要度において差が生じている場合、そのフレームの重要度に合わせて音声の音質を変化させることが可能となるので、視覚効果の高い動画及び音声の再生が可能になる。

【図面の簡単な説明】

【図1】

本発明で用いる動画像データの説明図である。

【図 2】

本発明の第 1 の実施の形態における画像処理装置 200 のブロック図である。

【図 3】

ROI 時間と非 ROI 時間の説明図である。

【図 4】

本発明の第 1 の実施形態における画像処理装置 200 が行う動画像データの符号化処理のフローチャートである。

【図 5】

動画像符号化データの構成の説明図である。

【図 6】

動画像符号化データの構成の説明図である。

【図 7】

ROI 設定部 208 のブロック図である。

【図 8】

ROI 設定部 208 の処理のフローチャートである。

【図 9】

デジタルカメラ（デジタルビデオカメラ）において ROI オブジェクトを選択する方法の説明図である。

【図 10】

ROI と ROI オブジェクトの説明図である。

【図 11】

ROI マスクの説明図である。

【図 12】

フレームデータ符号化部 202 のブロック図である。

【図 13】

フレームデータ符号化部 202 の処理のフローチャートである。

【図 14】

1 次元離散ウェーブレット変換の説明図である。

【図 15】

(a) は 4 つのサブバンドに分解する図、(b) は (a) の L L サブバンドを更に 4 つのサブバンドに分解する図、(c) は (b) の L L サブバンドを更に 4 つのサブバンドに分解する図である。

【図 1 6】

量子化ステップの説明図である。

【図 1 7】

(a)、(b)、(c) は各サブバンドにおける R O I と非 R O I の説明図である。

【図 1 8】

コードブロック分割の説明図である。

【図 1 9】

ビットプレーン分割の説明図である。

【図 2 0】

コーディングパスの説明図である。

【図 2 1】

レイヤー生成の説明図である。

【図 2 2】

レイヤー生成の説明図である。

【図 2 3】

フレーム符号化データの構成の説明図である。

【図 2 4】

音声データ符号化部 2 0 4 のブロック図である。

【図 2 5】

音声データ符号化部 2 0 4 の処理のフローチャートである。

【図 2 6】

動的ビット割り当て部 2 4 0 6 の処理のフローチャートである。

【図 2 7】

音声符号化データの構成の説明図である。

【図 2 8】

本発明の第3の実施の形態における画像処理装置2800のブロック図である。

【図29】

音声データ符号化部2801のブロック図である。

【図30】

最適音声符号化データ決定部2904のブロック図である。

【図31】

本発明の第2の実施の形態における画像処理装置3100のブロック図である。

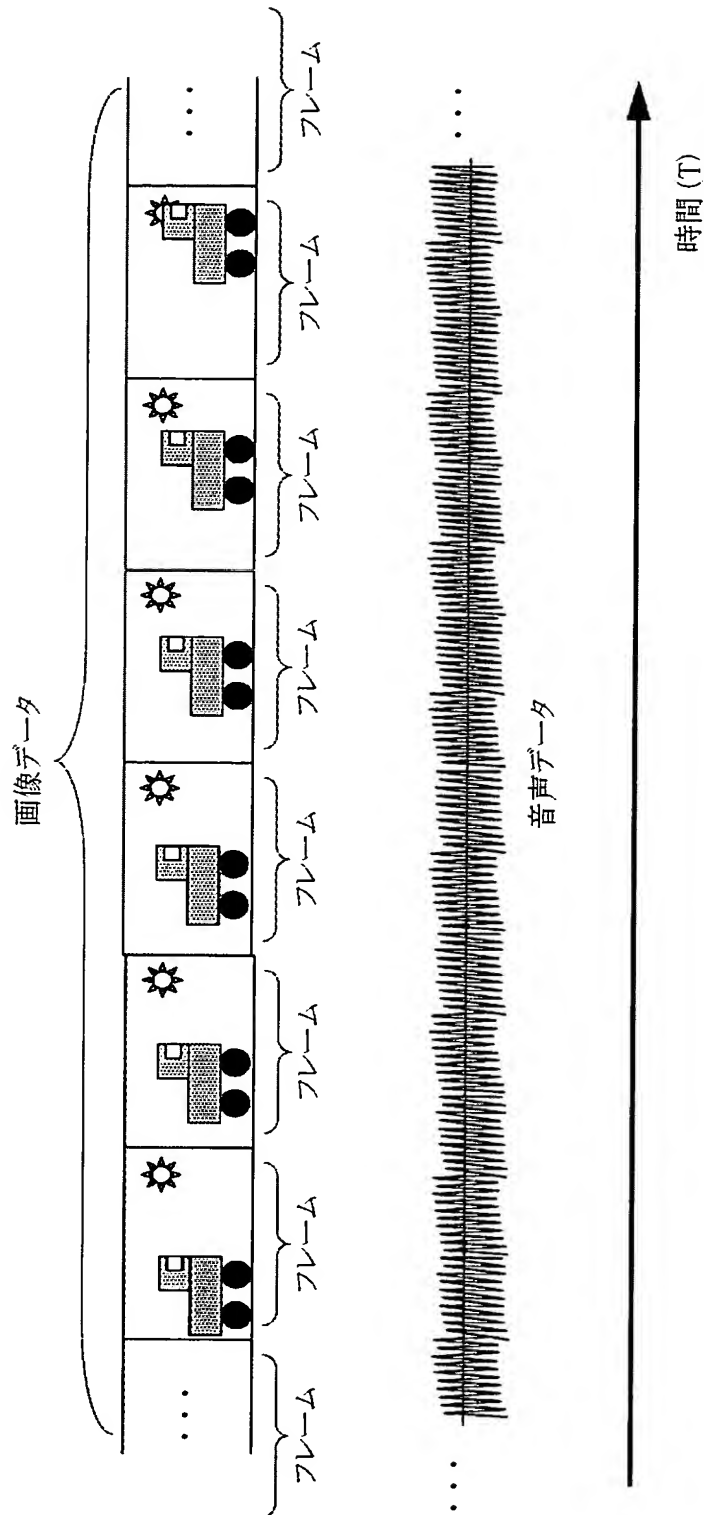
【符号の説明】

- 201 画像データ取得部
- 202 フレームデータ符号化部
- 203 音声データ取得部
- 204 音声データ符号化部
- 205 データ統合部
- 206 動画像符号化データ出力部
- 207 ROI設定指示部
- 208 ROI設定部
- 2801 音声データ符号化部
- 3101 スイッチ

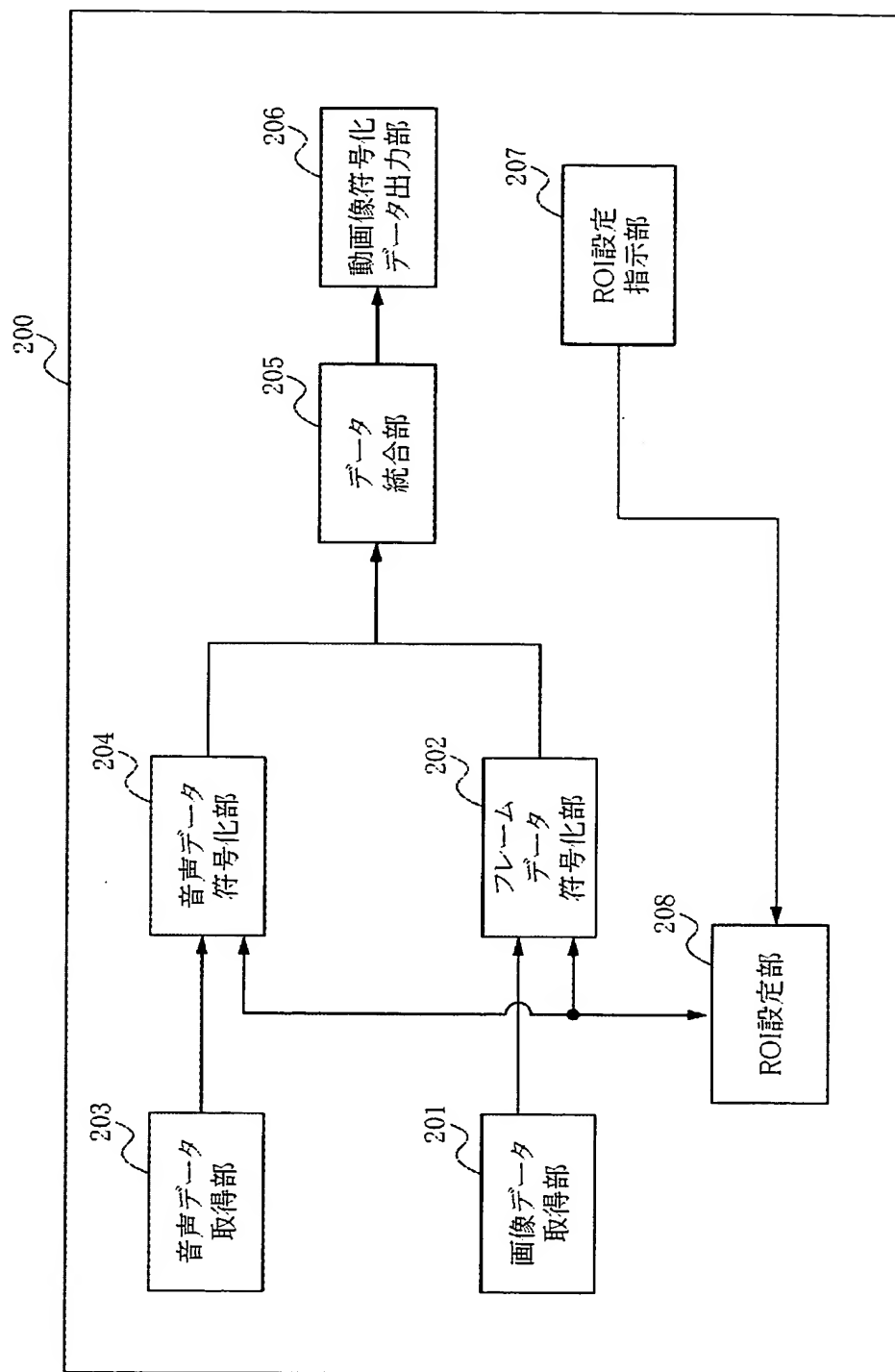
【書類名】

図面

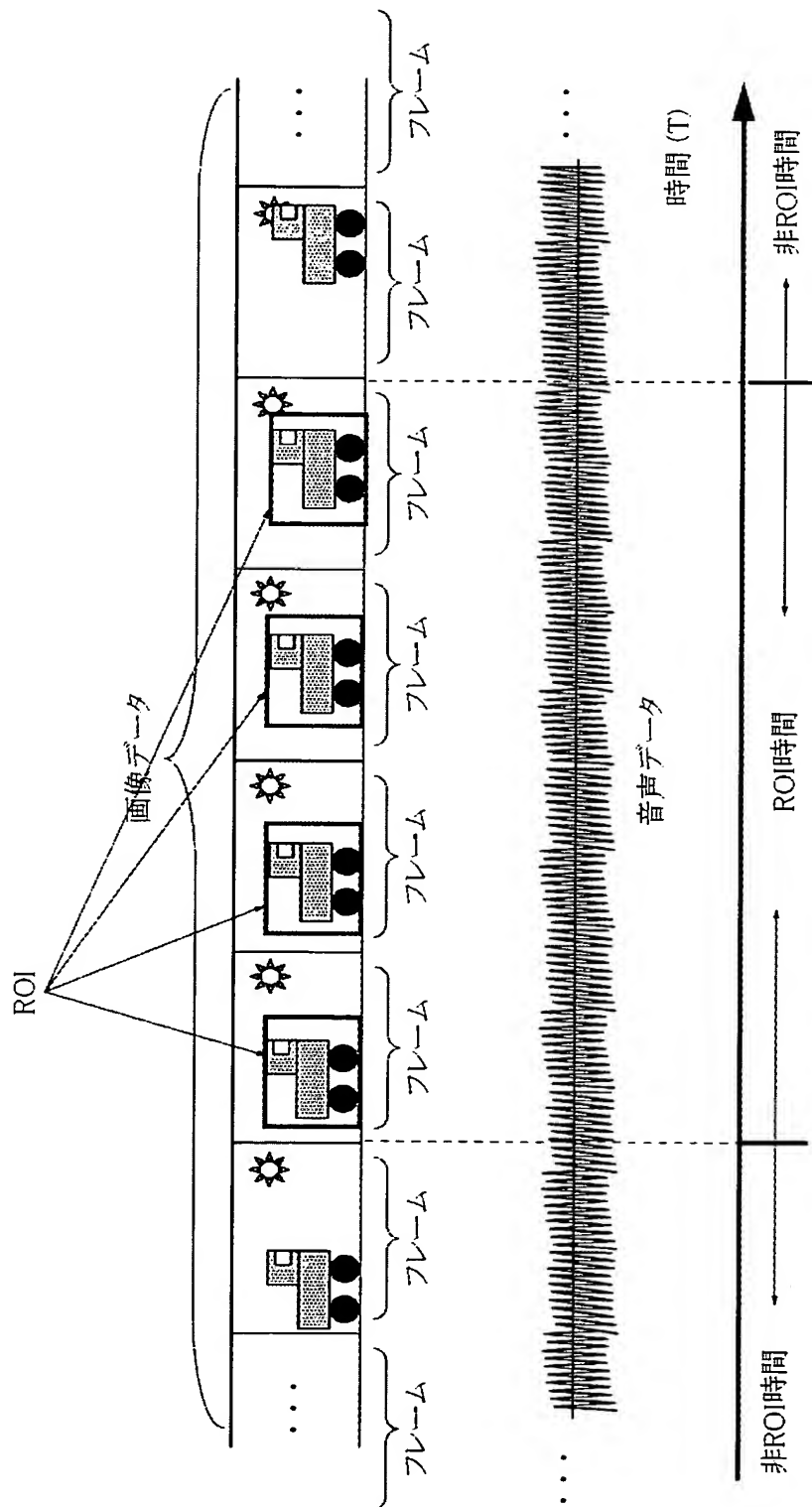
【図 1】



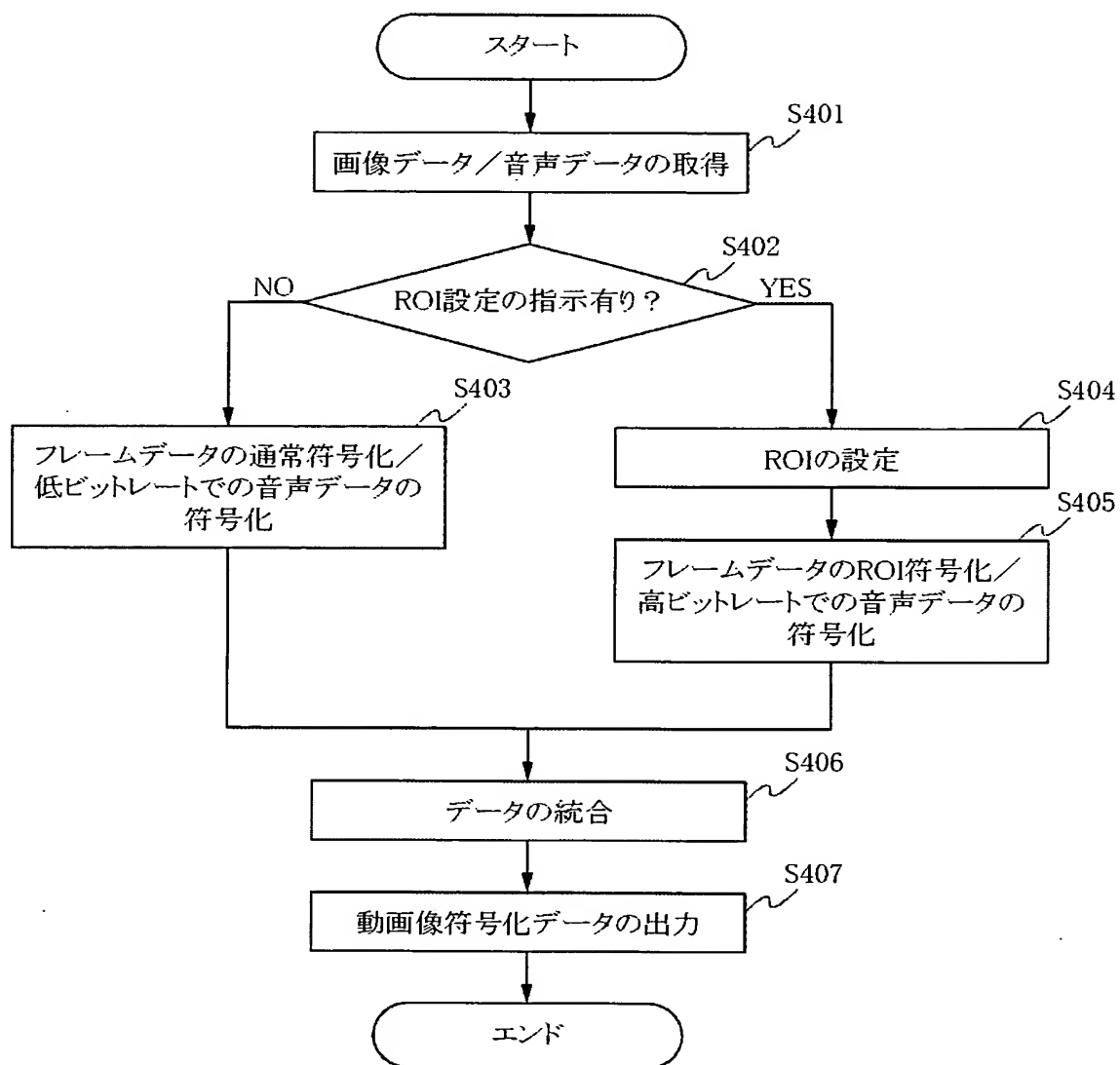
【図 2】



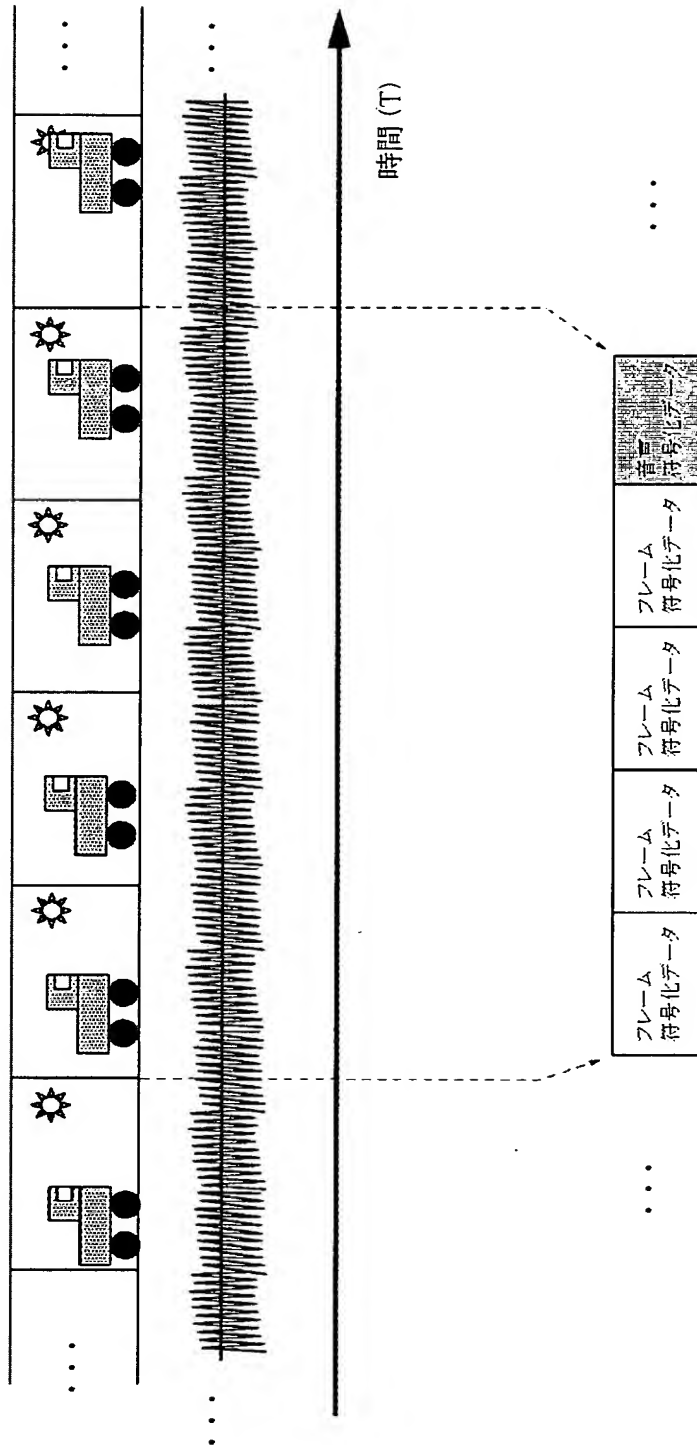
【図 3】



【図 4】



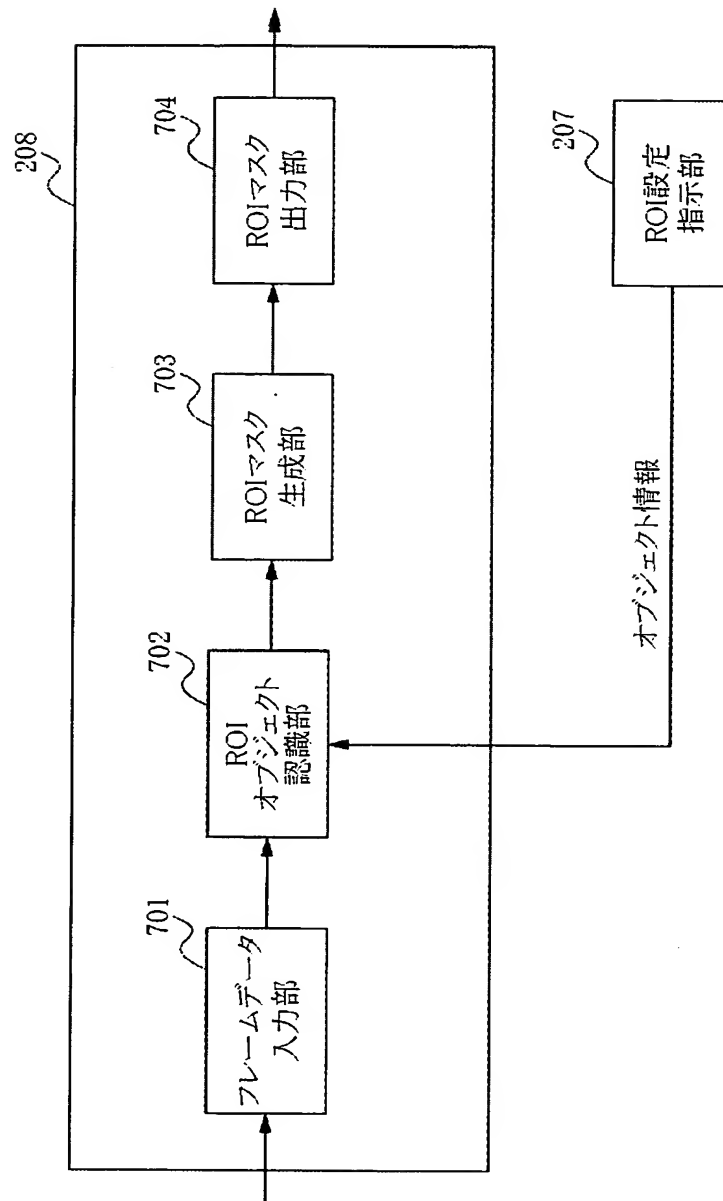
【図 5】



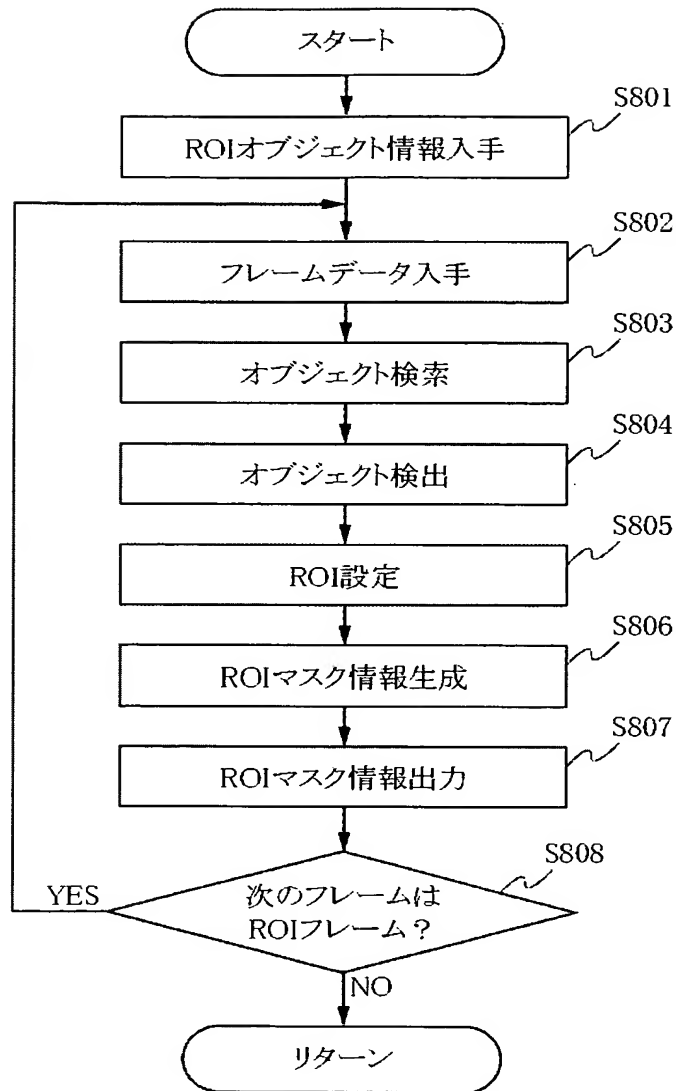
【図 6】



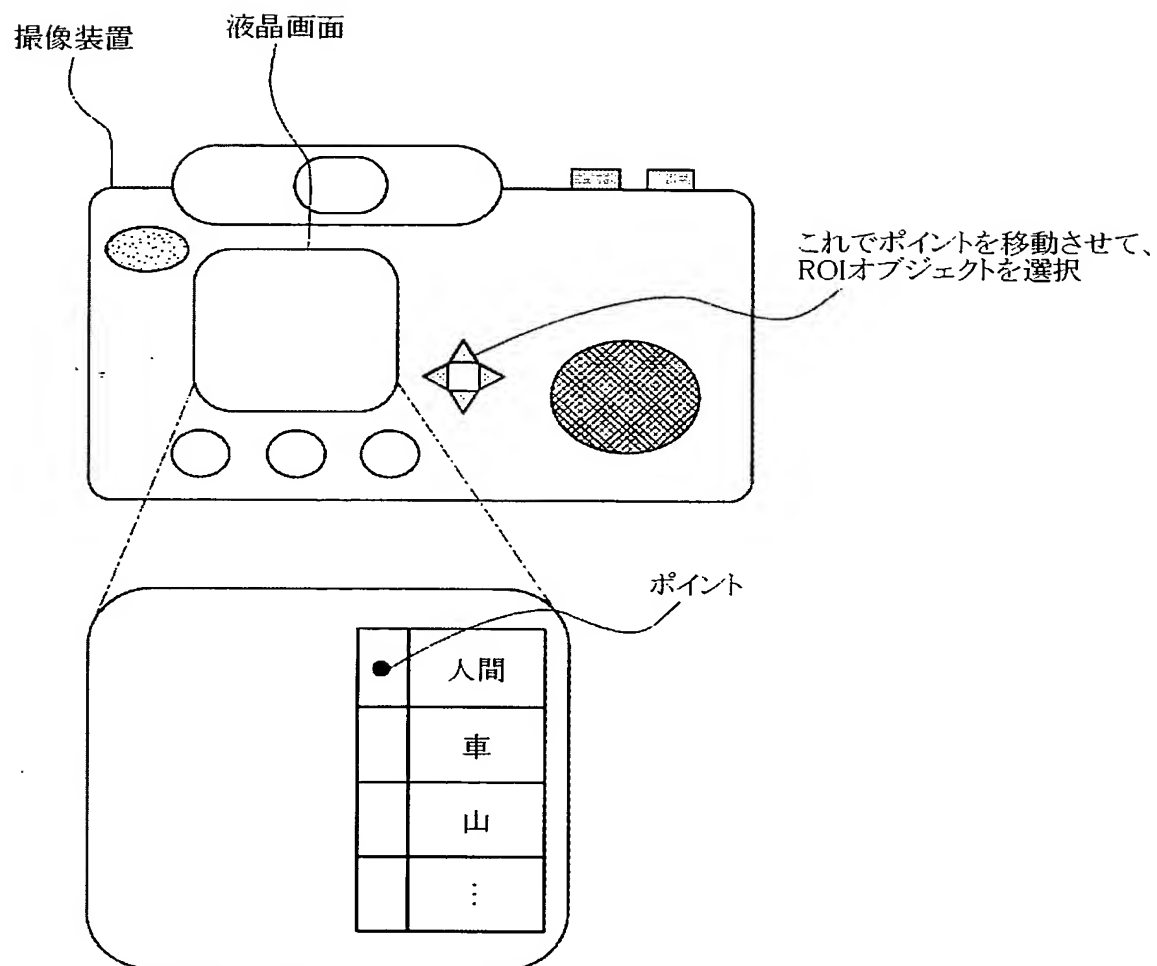
【図 7】



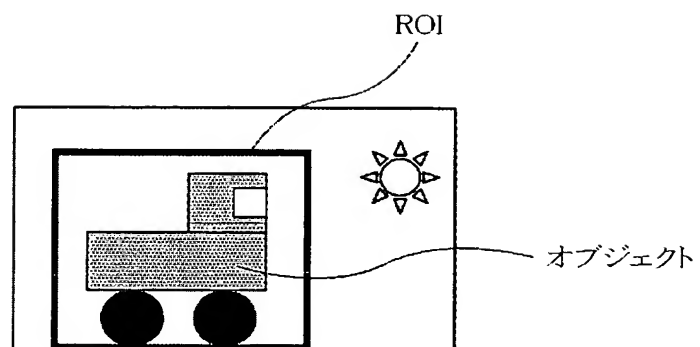
【図 8】



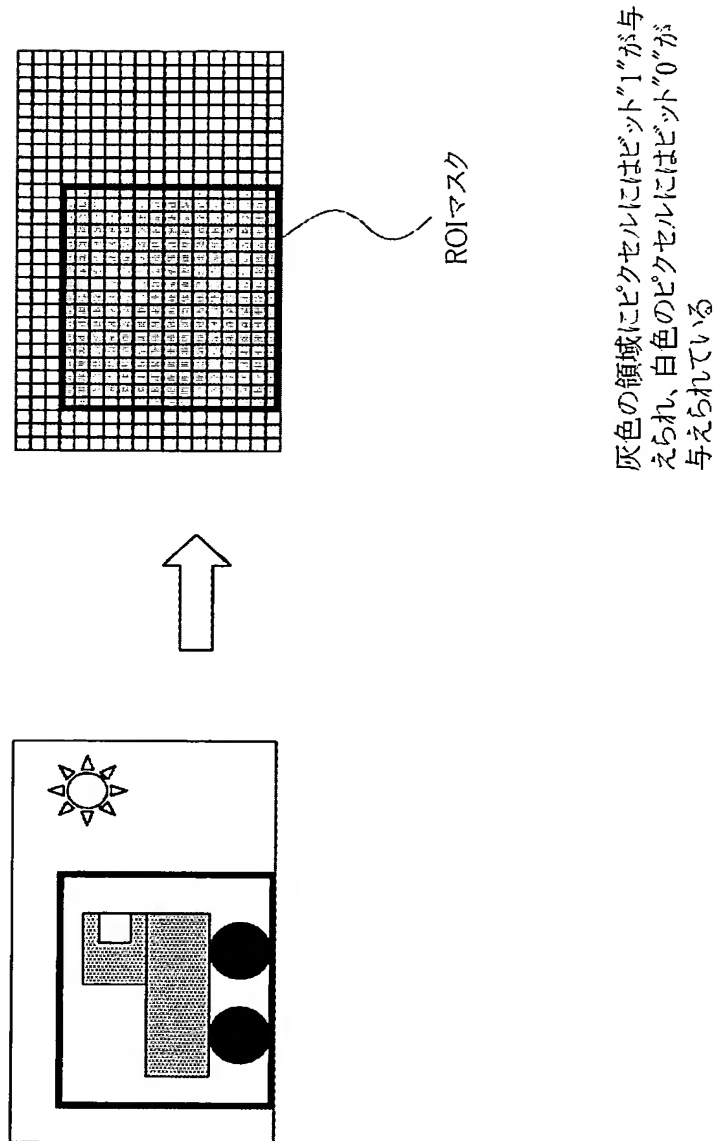
【図 9】



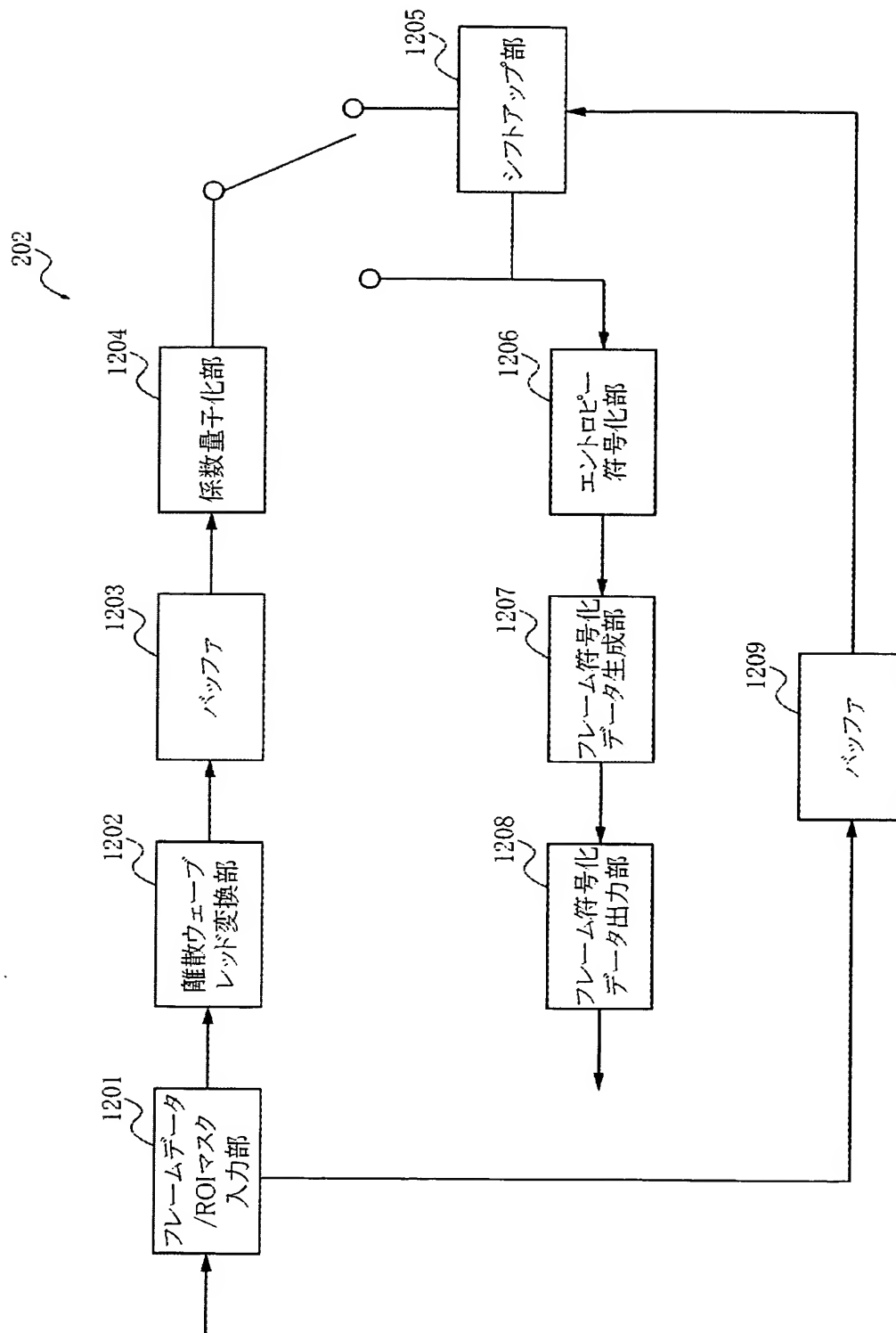
【図 10】



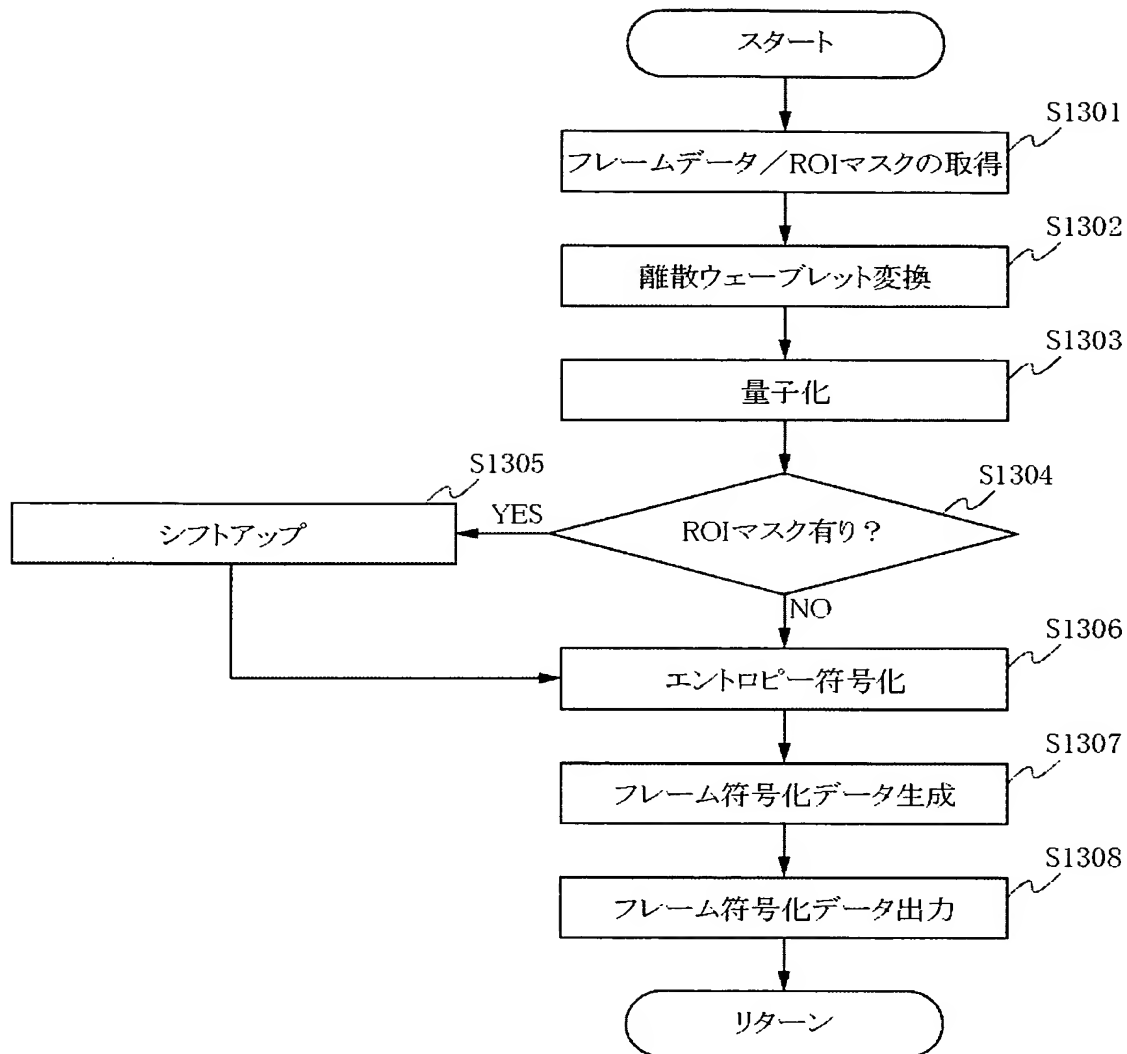
【図 11】



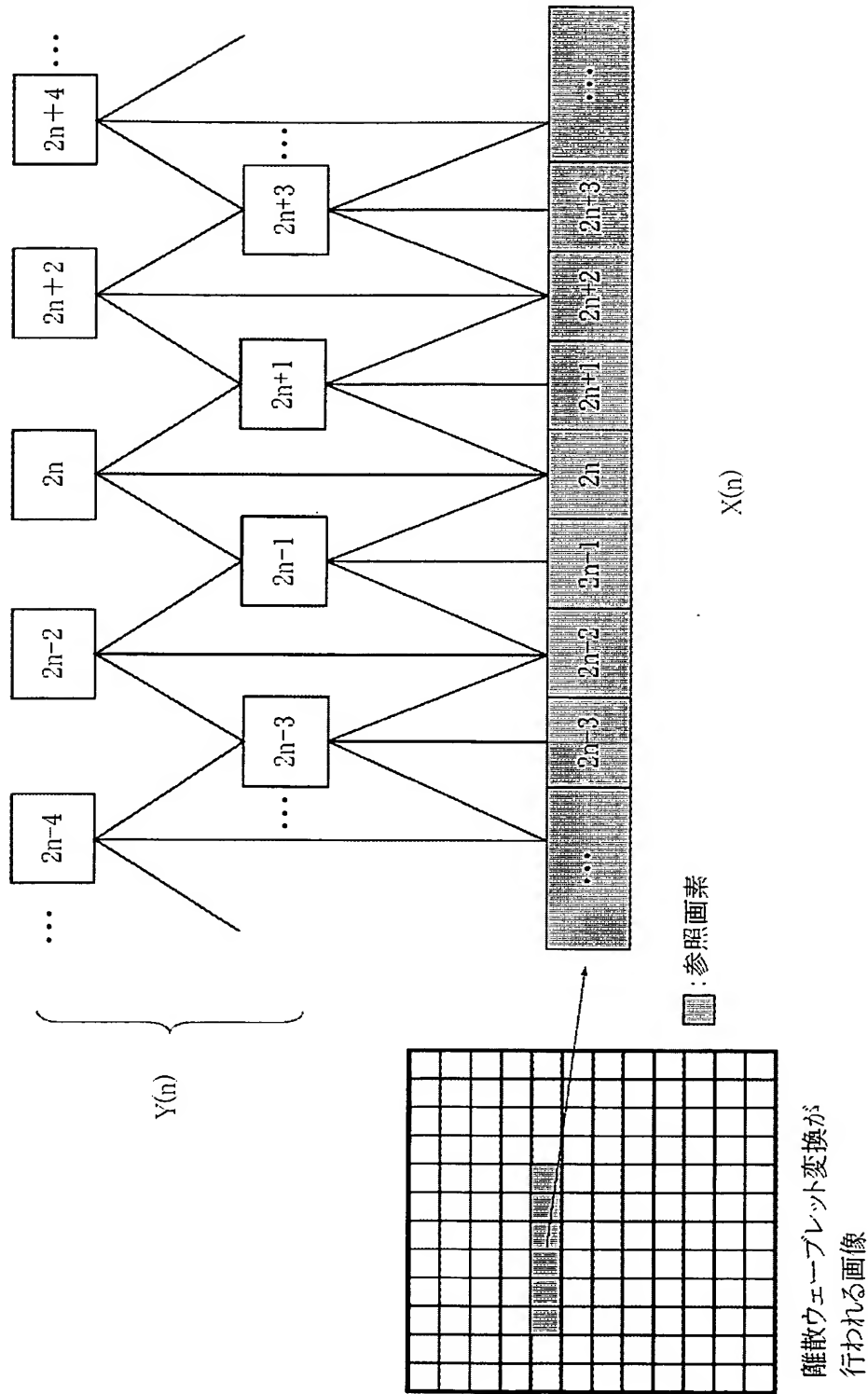
【図 12】



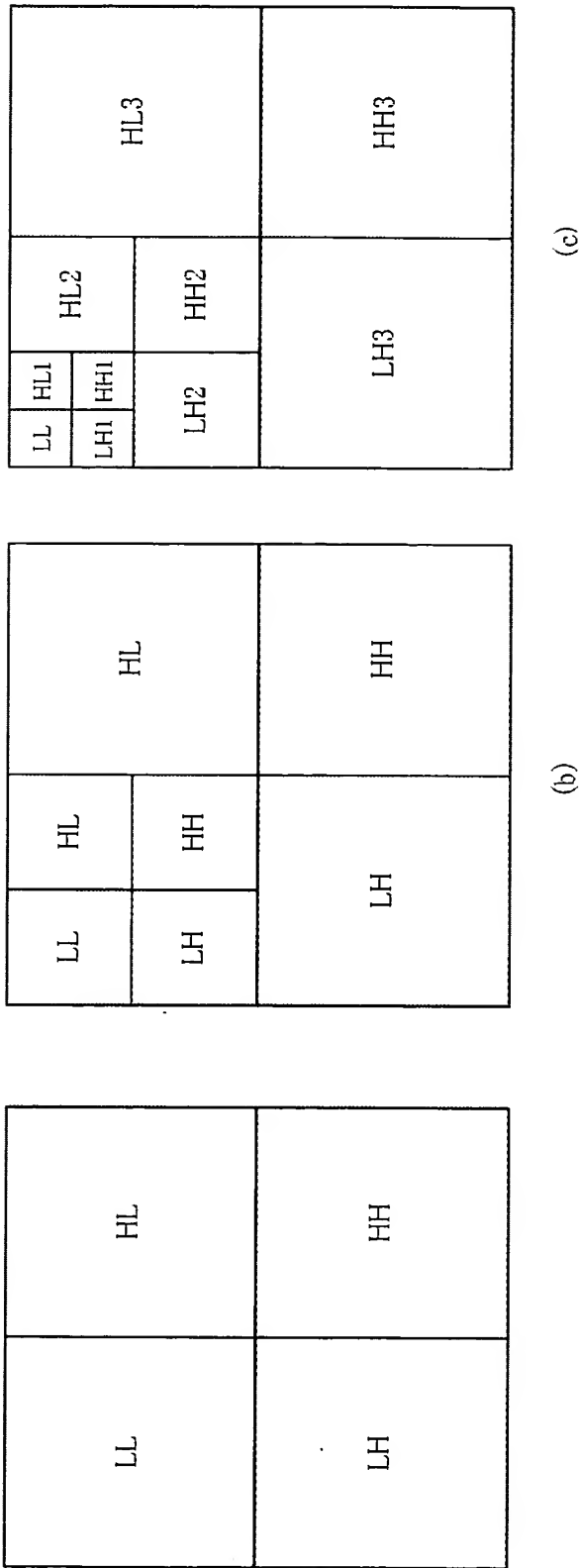
【図 13】



【図 14】



【図 1 5】

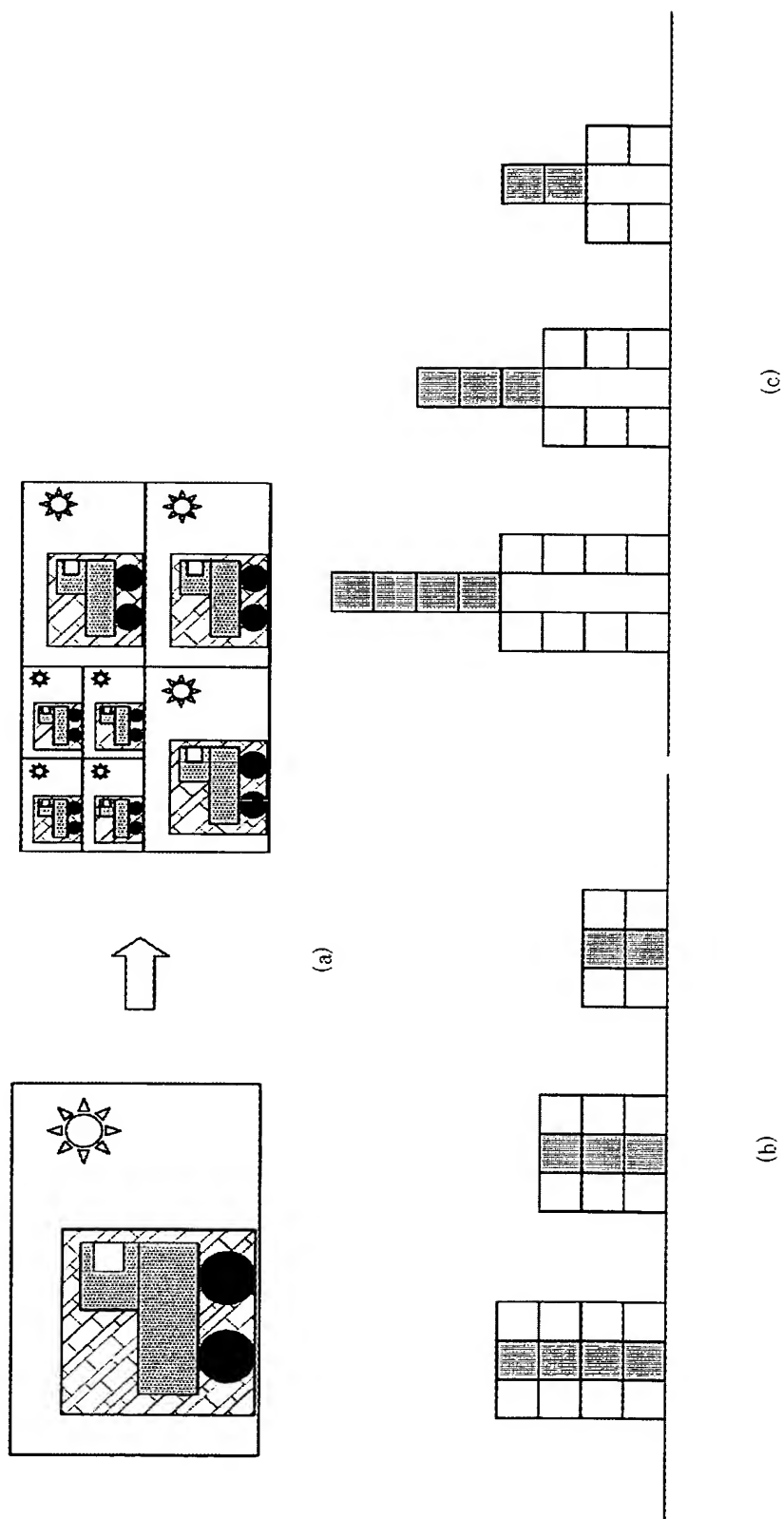


レベル0:LL, レベル1:HL1, HH1, LH1
レベル2:HL2, HH2, LH2, レベル3:HL3, HH3, LH3

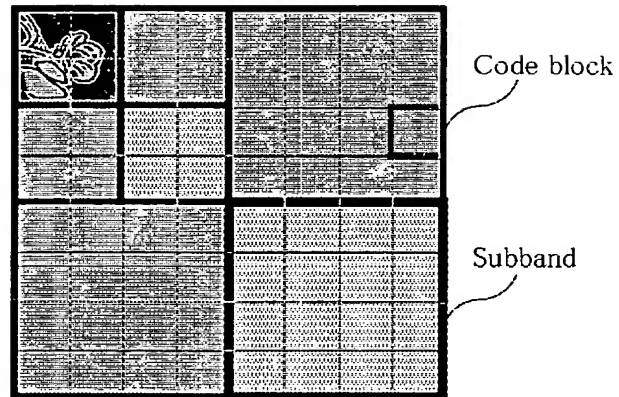
【図 1 6】

周波数成分	量子化ステップ
LL	1
HL1	2
HH1	2
LH1	2
HL2	4
HH2	4
LH2	4
HL3	8
HH3	8
LH3	8

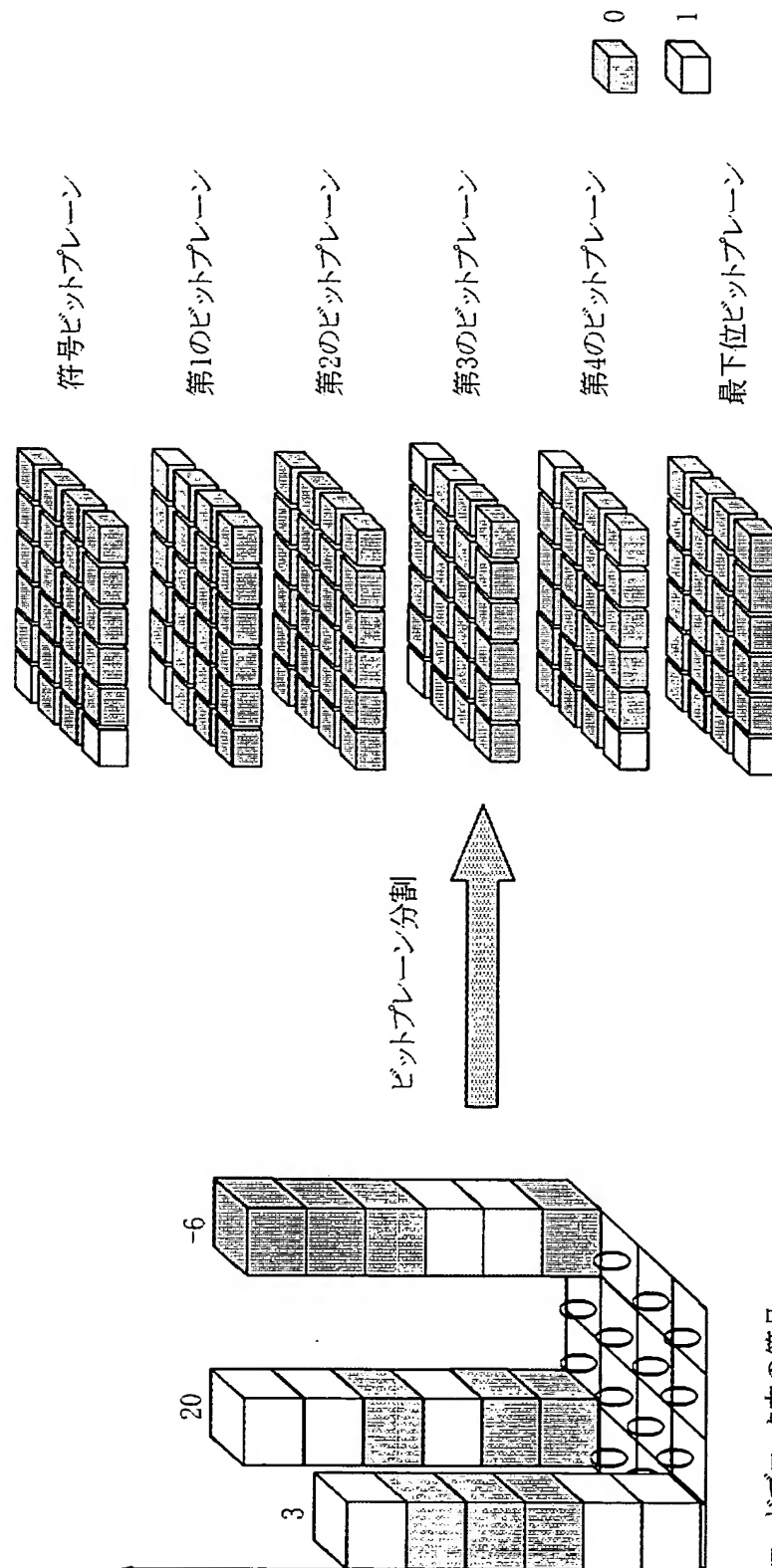
【図 17】



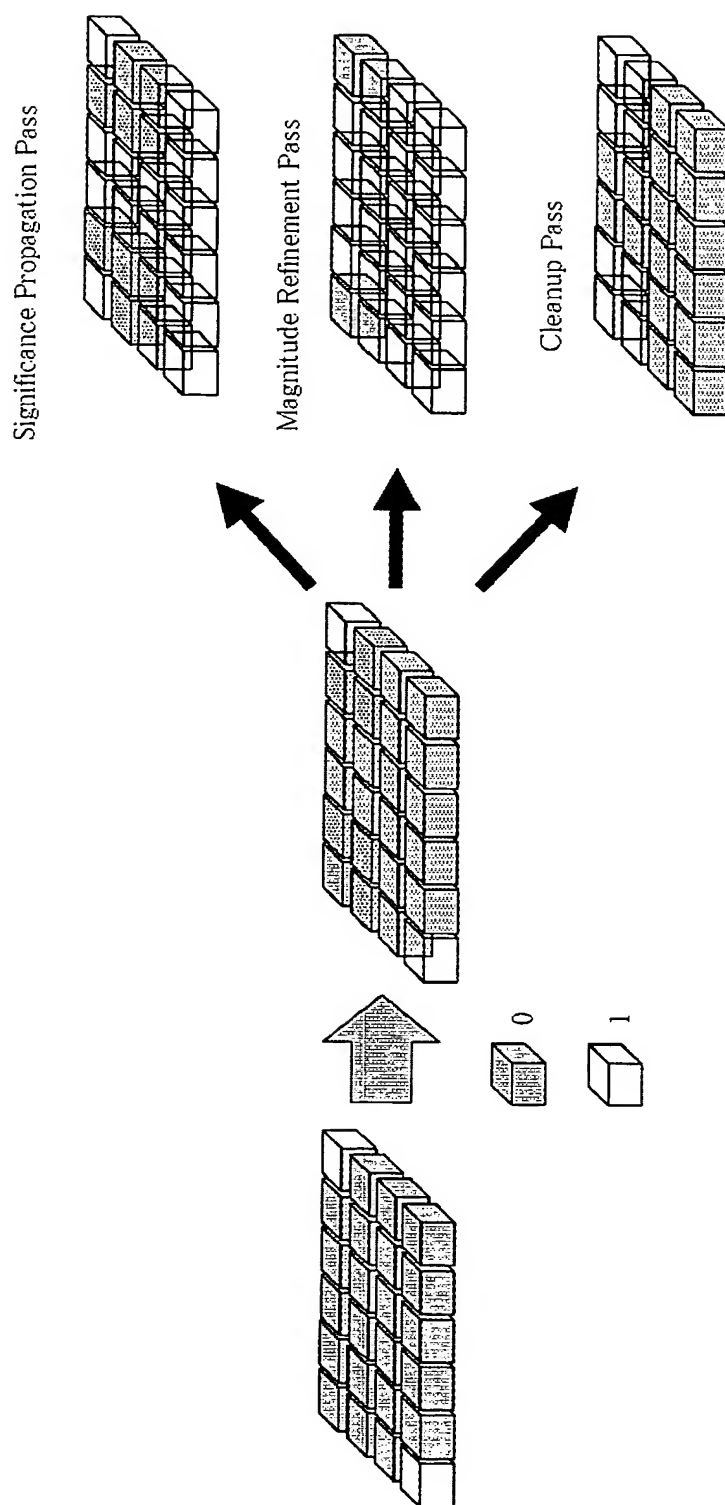
【図 18】



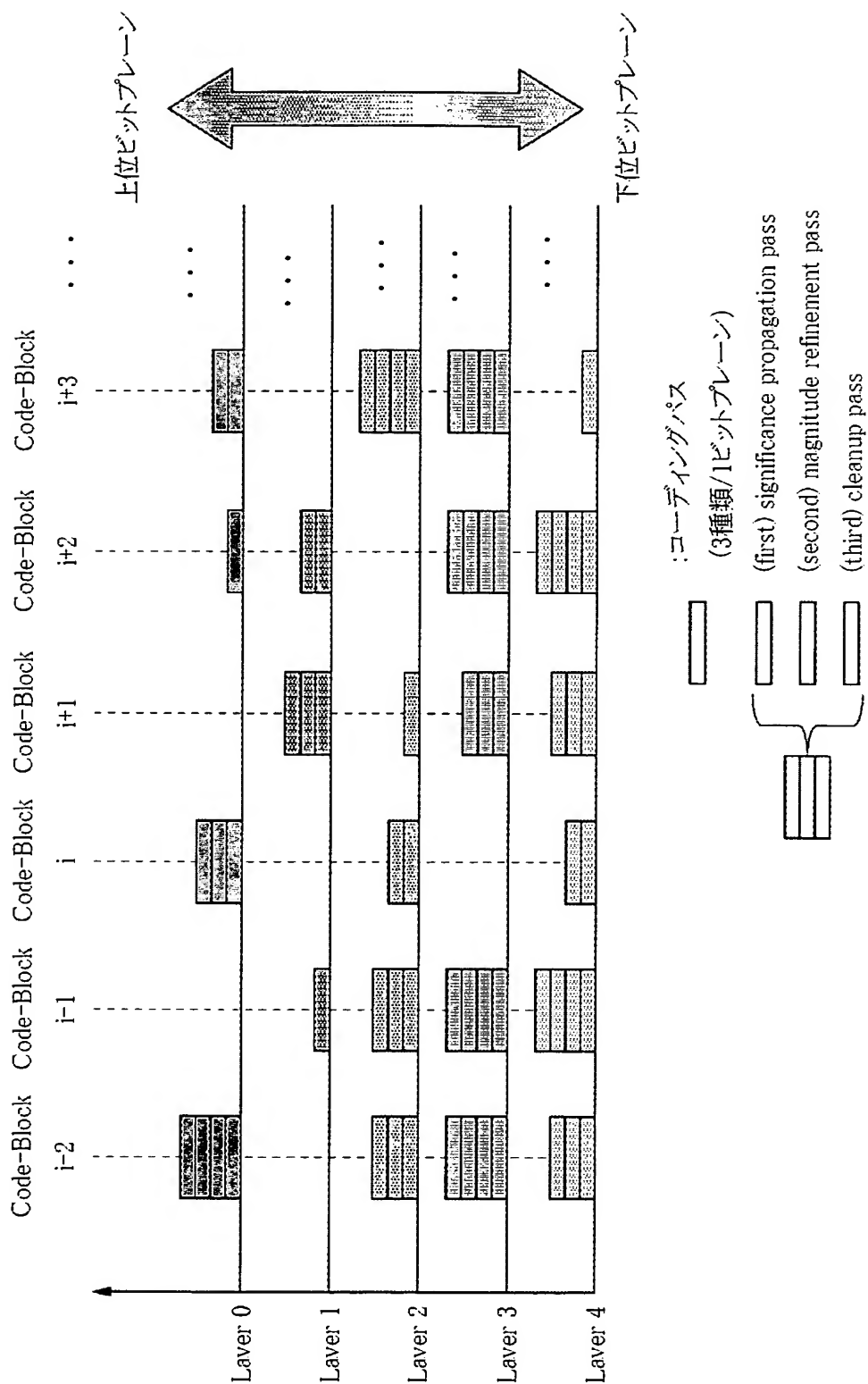
【図 19】



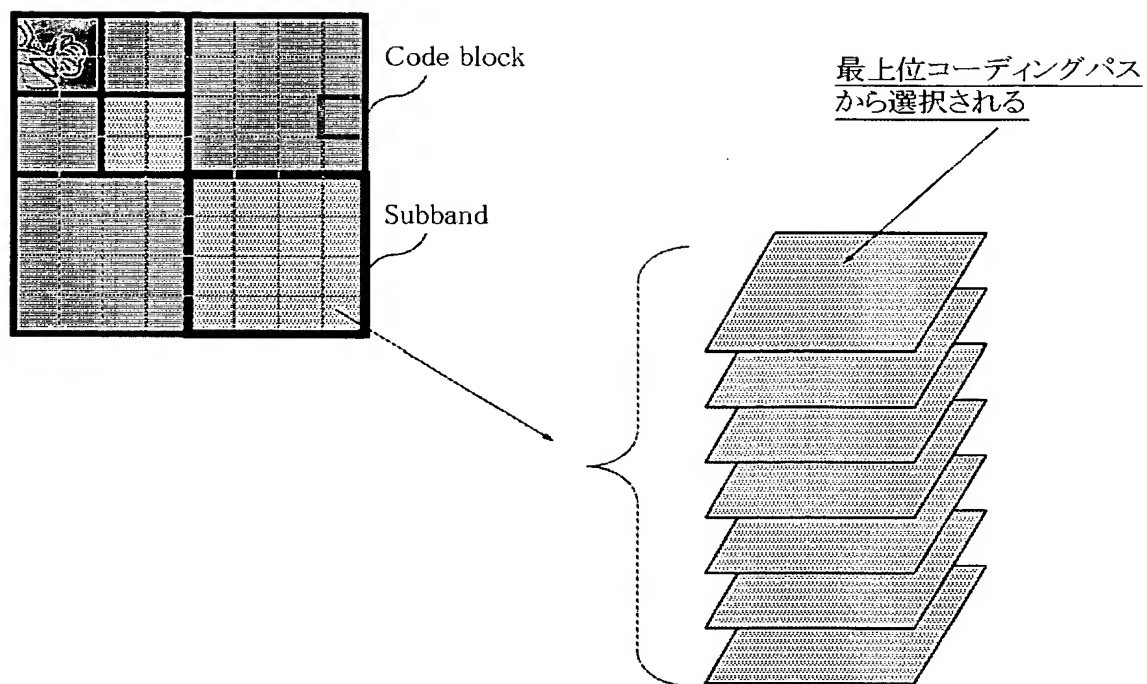
【図 20】



【図 21】



【図 22】

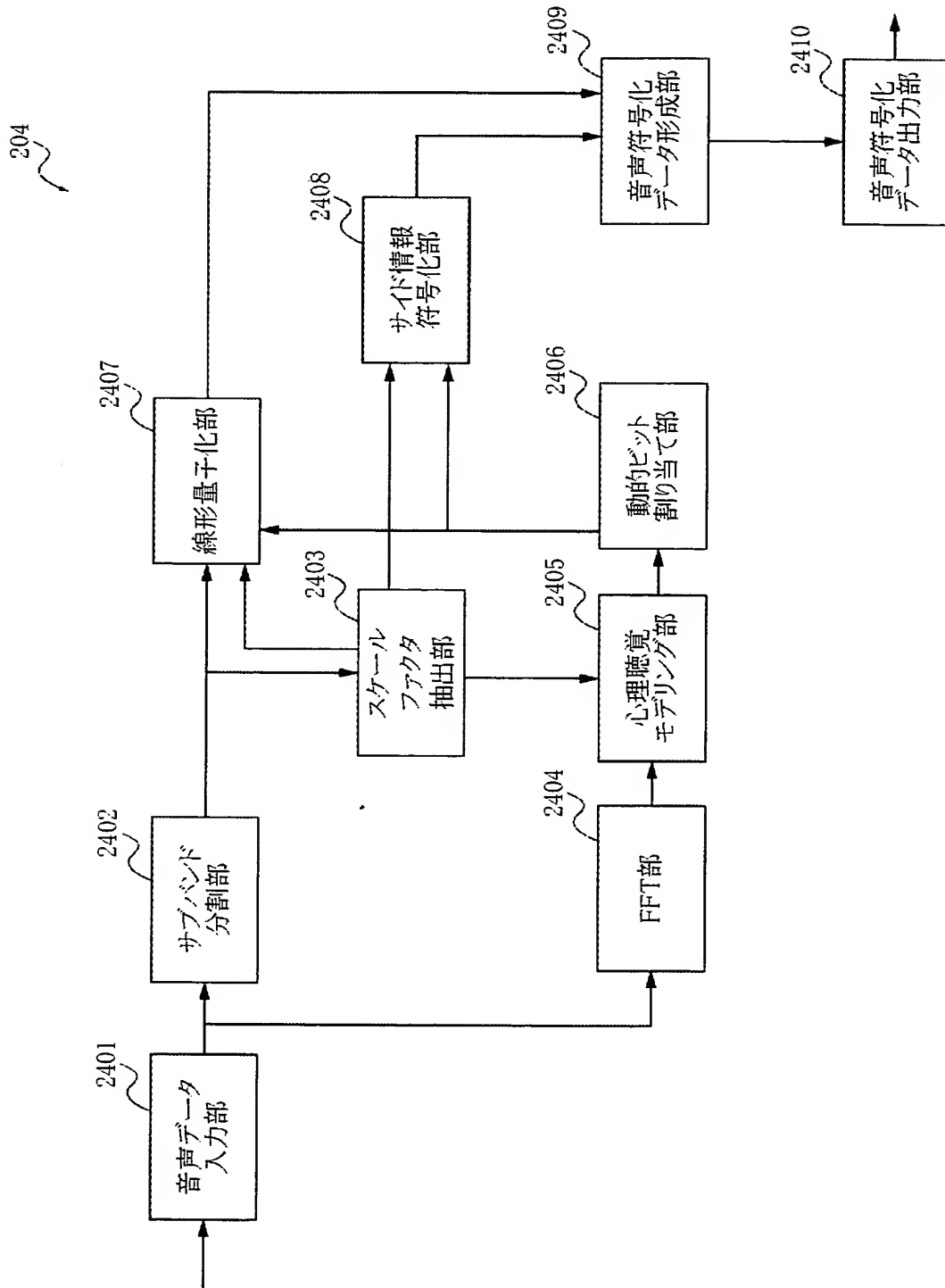


【図 2 3】

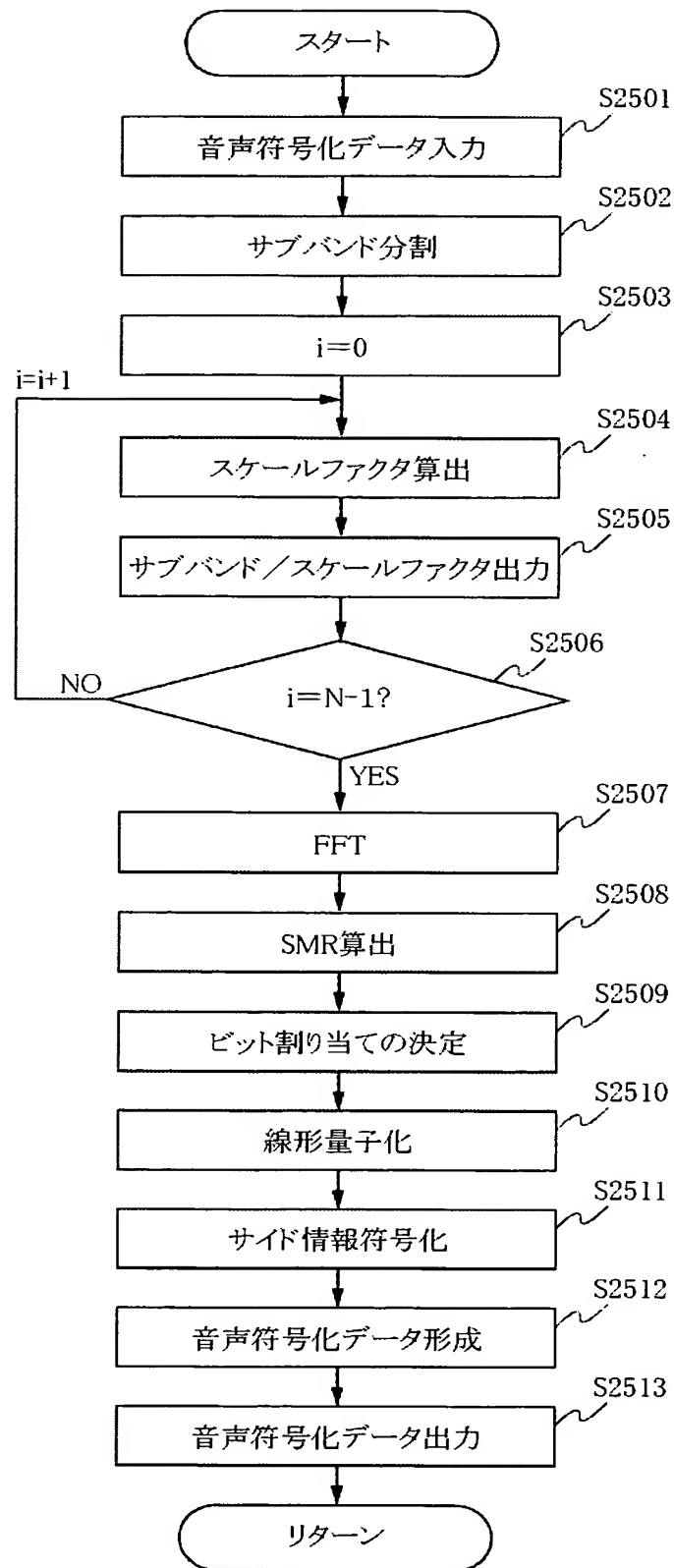
フレーム符号化データ

ヘッダ	レイヤー0	レイヤー1	レイヤー2	レイヤー3	レイヤー4
-----	-------	-------	-------	-------	-------

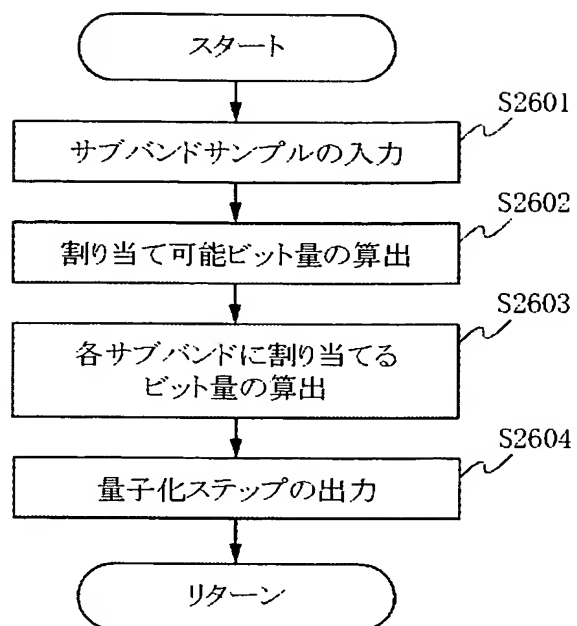
【図 24】



【図 25】



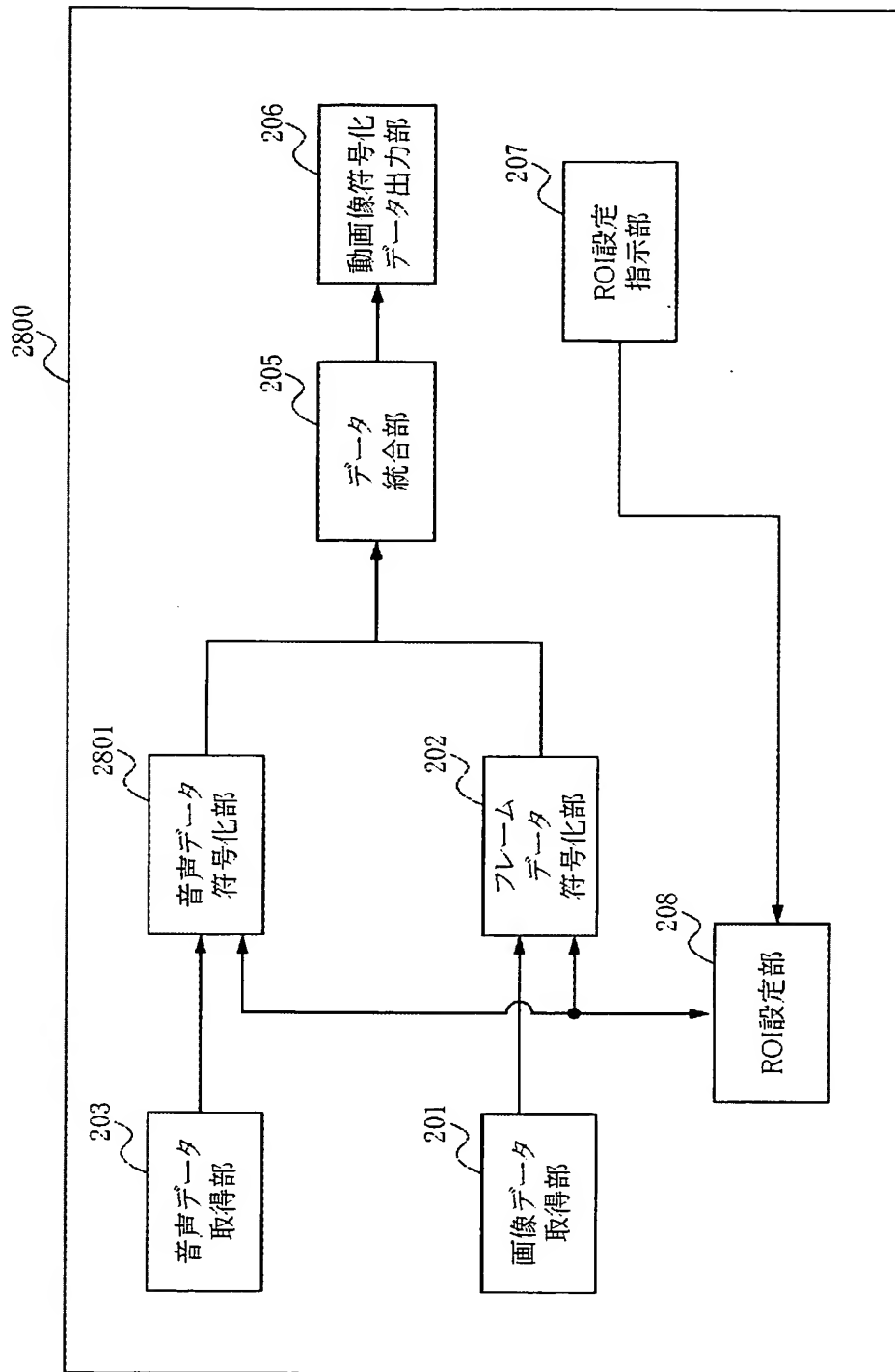
【図 26】



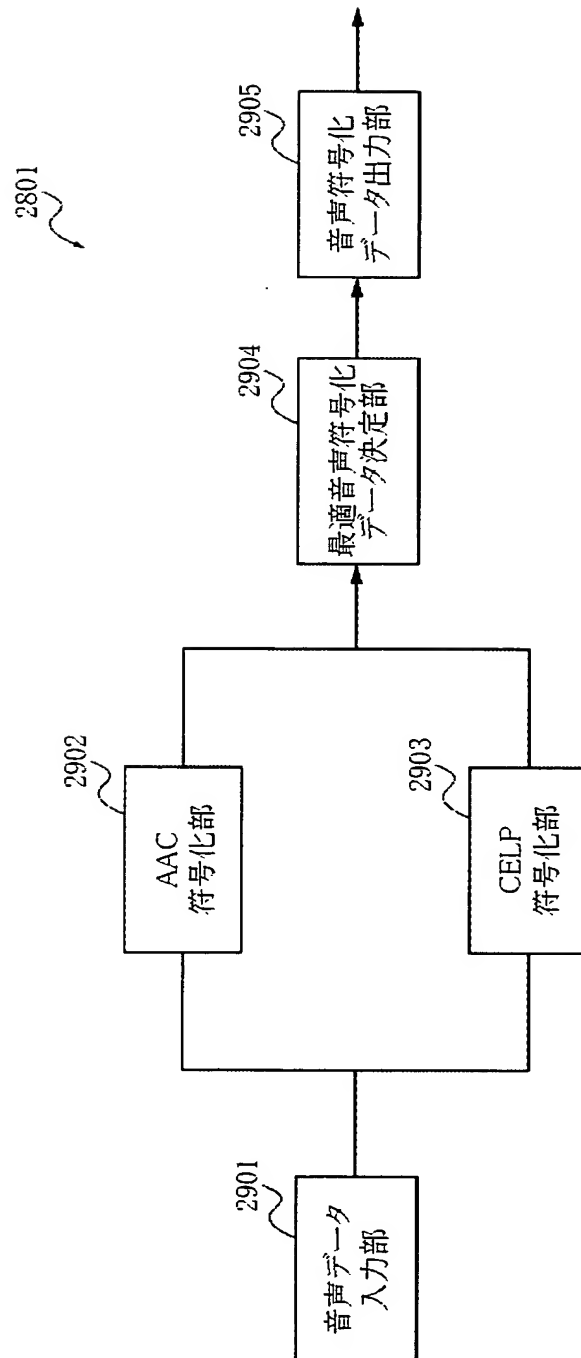
【図 2 7】

ヘッダ	サブバンドサンプル	アンシラリ データ
-----	-----------	--------------

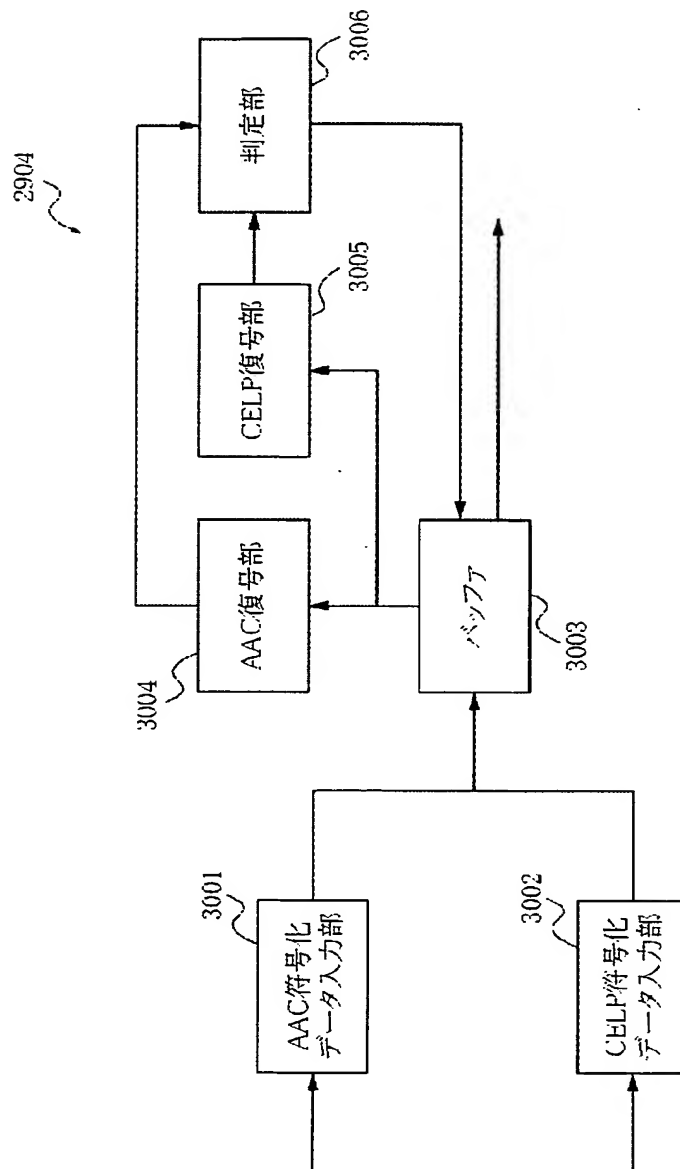
【図 28】



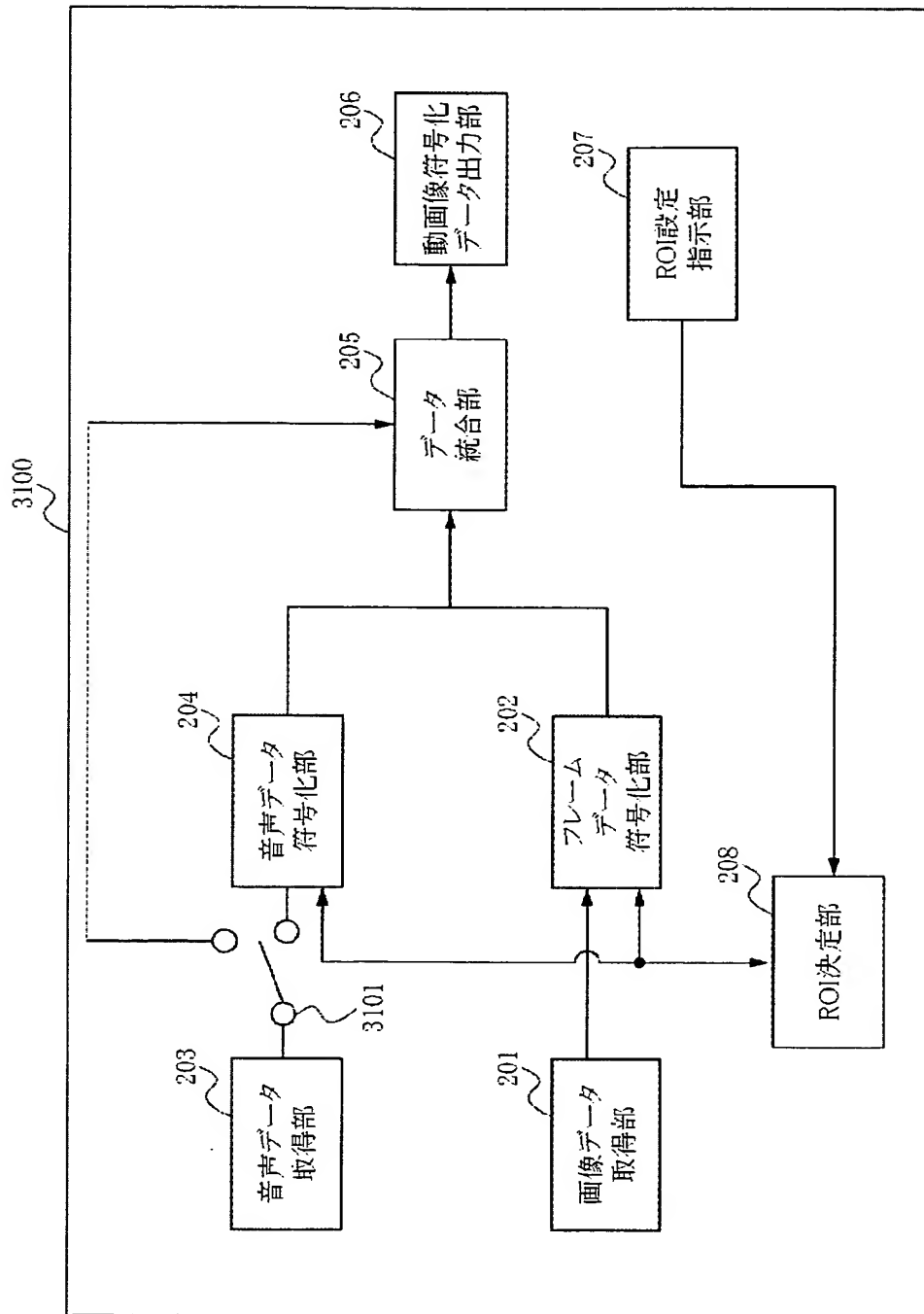
【図 29】



【図 30】



【図 31】



【書類名】 要約書

【要約】

【課題】 画像の重要度に応じて音声の音質を変更できるよう設定する画像処理装置を提供することを目的とする。

【解決手段】 入力した画像データを符号化する画像符号化部と、画像データとともに入力した音声データを符号化する音声符号化部と、画像の一部領域を高画質に符号化するように設定する重要画像の設定部と、重要画像の設定部における設定に従って、音声データを高品位に処理するよう設定する音声符号化設定部とを備える。

【選択図】 図 2

特願 2 0 0 2 - 2 7 3 2 5 4

出 願 人 履 歴 情 報

識別番号

[0 0 0 0 0 1 0 0 7]

1. 変更年月日

1 9 9 0 年 8 月 3 0 日

[変更理由]

新規登録

住 所

東京都大田区下丸子 3 丁目 3 0 番 2 号

氏 名

キャノン株式会社